THE UNIVERSITY *of* EDINBURGH

Structure Formation Within the Cosmic Web

Elizabeth Eardley

A thesis submitted in fulfilment of the requirements
for the degree of Doctor of Philosophy
to the
University of Edinburgh
2015

Abstract

In this era of high-precision cosmology we are able to measure and predict properties of the large-scale structure of our Universe to a fine degree. However we still lack a clear and tested understanding of the effects of the large-scale environments on galaxies and their host halos. This thesis focuses on bettering our understanding of this issue by investigating the dependence of galaxies and halos on their location within the cosmic web.

An algorithm based on the tidal tensor prescription is developed and applied to the MDR1 1 $(h^{-1}\mathrm{Gpc})^3$ dark matter simulation to classify the geometric environment of every location in the simulated volume as one of the four components of the cosmic web; voids, sheets, filaments and knots. Conditional halo mass functions are extracted to investigate the influence of tidal forces on the abundances and mass distribution of dark matter halos. A theoretical framework based on Gaussian statistics is presented and used to derive predictions for halo abundances in different geometric environments. The Gaussian theory predicts no coupling of tidal forces and, hence, that the halo mass function is independent of geometric environment for a given local mass density. It is shown that the halo mass functions extracted from the simulation are fully consistent with this picture.

It is then shown how this method of classifying geometric environments can be extended to observational datasets. The Galaxy And Mass Assembly (GAMA) spectroscopic redshift survey, with its wide field and high completeness, is excellently suited to this study. The geometric environments of the three equatorial GAMA fields are classified, following a thorough analysis of the additional uncertainties introduced when moving to observational datasets. Additionally, the geometric environments of the GAMA galaxies and groups are classified, allowing the influence of the cosmic web on large-scale structure to be

investigated.

Both the galaxy luminosity function and the group mass function within the observed cosmic web are studied and no evidence of a direct impact of the web is seen. It is found that all modulations can be fully attributed to the indirect dependence of these properties on the local matter overdensity. Whilst these results indicate that there is no strong dependence of the scalar properties of large-scale structure on geometric environment, the final investigation of this thesis presents an attempt to look in more detail at the environmental dependence of stellar properties by investigating stellar-formation histories within the cosmic web.

Declaration

Except where otherwise stated, the research undertaken in this thesis was the unaided work of the author. Where the work was done in collaboration with others, a significant contribution was made by the author, as summarised below:

- Chapter 2 is based on the published work of
  Alonso D., Eardley E., Peacock J. A., 2015, MNRAS, 447, 2683.
  All analysis of simulated data was undertaken solely by the author. A significant contribution to the text and all figures in this chapter were produced by the author. The analytic calculations were produced by D. Alonso and are reproduced here to allow the comparison of theory and data.

- Chapter 3 and Section 4.1 include work published in
  Eardley E. et al. MNRAS, 448, 3665,
  a paper produced in conjunction with members of the GAMA team, and of which the author was lead author. The luminosity function calculations of Section 4.1 were supplied by Tamsyn McNaught-Roberts.

- Chapter 5 presents initial results of a project conducted in collaboration with Rita Tojeiro. All text and main figures were produced by the author.

E. Eardley
March 2015

Acknowledgements

First and foremost I would like to thank my supervisor, Prof. John Peacock, without whom this thesis would not have been possible. I am grateful for the knowledge and direction he provided throughout my PhD along with his kindness, patience and humour.

I also thank Dr. Catherine Heymans, my second supervisor, for her support and encouragement, for always being approachable, enthusiastic and a valuable source of practical advice.

Of the many people who contributed to my development as a cosmologist I offer special thanks to Rita Tojeiro, David Alonso, Peder Norberg and Tamsyn McNaught-Roberts for their contributions to this thesis.

I thank my fellow PhD students and all the members of the ROE for creating a friendly work environment, and Elliot Marsden and Fergus Cullen for many tea breaks in a friendly home environment.

Finally, I thank my parents for their support and for always being there.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

Cosmology is the scientific study of the large-scale properties of the Universe: its very beginnings, its dynamics and constituents, and its ultimate fate. Since Einstein's theories of relativity transformed the subject a century ago, a standard model of cosmology has risen to wide spread acceptance as an accurate description of the Universe. The standard model provides us with previously unthought of predictions regarding the nature of space and time, which have passed many tests to date. Today we have the ability to test this widely assumed model of cosmology with unprecedented scope and accuracy.

The standard model relies on general relativity as a geometric theory of gravitation to describe the nature of gravity on all scales. It draws on the cosmological principle, which states that there is no special point or direction in the Universe, and thus, on large enough scales, the Universe is spatially homogeneous and isotropic. The resulting picture is a Universe which originated from a singularity around 13.8 billion years ago, and has undergone cosmic expansion ever since. The standard model is also referred to as the ΛCDM model for its predictions on the constituents of our Universe. Λ represents the cosmological constant which parameterises the dark energy of the Universe, and CDM represents the cold dark matter thought to be the most abundant, but weakly interacting, type of matter. Although these components of the standard model mathematically explain many observations, little is known about their forms. This leaves us with the disconcerting picture that only $\sim 5\%$ of the mass-energy content is in the form of the relatively well understood 'ordinary' matter.

Much of the work in modern cosmology has focused on testing this standard

model in great detail and searching for clues on the nature of the poorly understood constituents of the Universe. Recent developments in the field of observational cosmology have provided expansive and detailed datasets which map out the distribution of matter, allowing insights into many aspects of cosmology and the formation and evolution of structure. Similarly, with the advent of large numerical simulations we are now able to simulate a picture of a universe with a known cosmology and compare it to our own.

Whilst the Universe is understood to have originated from a very smooth, near-homogeneous state, today it is observed to contain a complex distribution of structures. These structures are a result of the gravitational collapse of matter. Such collapsed objects cover a vast range in scale, including planets such as our own and the sun it orbits, which is only one of billions of similar stars which form the galaxy in which we reside. The galaxy itself is again one of very many which have been observed and can themselves form larger structures - galaxy clusters. These clusters are understood to trace the dark matter within our Universe, which itself collapses under the forces of gravity to create dark matter halos. All these objects are not randomly positioned throughout space, but have evolved and moved over cosmic time due to gravitational forces.

Additionally, non-gravitational forces become increasingly important on sub-cluster scales. Whilst considering only gravitational forces results in a reasonable approximation, baryonic effects, such as radiative cooling and feedback (from supernovae and accreting super-massive black holes), can have a significant impact on structure formation (Duffy et al. 2010). However modelling these non-gravitational processes is an enormously complex and difficult problem, which remains an active field of study.

The resulting distribution of matter is observed to be a complex multi-dimensional web-like distribution known as the cosmic web. The study of large-scale structure (LSS) refers to the investigation of these structures. From such investigations we hope to understand how individual objects such as galaxies and dark matter halos form and evolve and to gain insights into the underlying nature of our Universe.

The Universe can thus be studied on a vast range of scales, all of which are necessary to understand the rules which govern it and how it came to produce the structures we observe. In order to study the distribution and properties of

collapsed objects in our Universe, we first need some theoretical models against which to compare observational data. The remainder of this chapter attempts to provide an overview of the theoretical background necessary to interpret and utilise the available data on the LSS contained in our Universe.

1.1 Cosmological background

1.1.1 Friedmann-Lemaître-Robertson-Walker metric and Friedmann's equations

Much of modern cosmology has developed from the belief that we, as observers on earth, are not positioned in any privileged or special location within the Universe. It follows then, that if the Universe is isotropic about our location it can be presumed to be isotropic about any location in the Universe. Many observational studies have found just that – the large-scale Universe is homogenous and isotropic (Wu et al. 1999, Yadav et al. 2005). Homogeneity requires the density of the Universe to be the same, when viewed on large enough scales, from any position in the Universe. Similarly, isotropy requires that the Universe appears the same when viewed in any direction. In other words, homogeneity and isotropy imply translational and rotational invariance respectively. These two properties form the basis of the widely assumed cosmological principle, which states that our Universe is both homogenous and isotropic on large scales.

On small scales, fluctuations around the background level break homogeneity. However, even on these scales the Universe can be thought of as statistically homogeneous and isotropic. That is, if it were possible to measure the density a number of times in an ensemble of universes, the average over this large set should obey translational and rotational invariance. Similarly, if we were able to move around in the Universe and repeat the density measurements from a large number of randomly selected locations, we would again expect the results to display homogeneity and isotropy on all scales.

An isotropic and homogenous Universe can be described within the framework of general relativity by the Friedmann-Lemaître-Robertson-Walker (FLRW) space-time metric. In this model, the space-time interval between two events

is expressed using spherical coordinates as

$$ds^2 = -c^2dt^2 + R^2(t)\left[\frac{\mathrm{d}r^2}{1-kr^2} + r^2(\mathrm{d}\theta^2 + \sin^2\theta \mathrm{d}\phi^2)\right]. \tag{1.1}$$

Here (r, θ, ϕ) are comoving coordinates, c is the speed of light, $R(t)$ is the scale factor describing the expansion of space and k is a dimensionless parameter which describes the curvature of the Universe. $k > 0$ describes a closed Universe with positive curvature, $k < 0$ describes an open Universe with negative curvature, and $k = 0$ describes a flat Universe with no curvature.

The curvature of the Universe alters the path that CMB photons will have taken to get to us, and hence information on the curvature of the Universe can be extracted from observations of the CMB; if the Universe were flat, the dominant angular scale of fluctuations in the microwave background would be expected to be about one degree. In open and closed universes this dominant scale would be expected to be smaller and greater, respectively. The size of the temperature fluctuations within the observed CMB indicate that our Universe is flat, and $k = 0$ is henceforth assumed (Miller et al. 1999, de Bernardis et al. 2000).

A dimensionless scale factor can be introduced by normalising $R(t)$ relative to its value today, $R(t_0)$:

$$a(t) = \frac{R(t)}{R(t_0)}. \tag{1.2}$$

The presence of this time-dependent scale factor implies that the Universe is expanding. This has been strongly verified by the observational evidence of a redshift-distance relation. Redshift refers to the change in the wavelength of light between emission, λ_{em}, and observation, λ_{ob},

$$z = \frac{\lambda_{\mathrm{ob}} - \lambda_{\mathrm{em}}}{\lambda_{\mathrm{em}}}. \tag{1.3}$$

Within the FLRW metric the wavelengths of light at two different epochs are directly related to the scale factors at those epochs. Hence, the redshift at a time t_{em} is given by

$$z = \frac{a(t_{\mathrm{ob}})}{a(t_{\mathrm{em}})} - 1 = \frac{\lambda_{\mathrm{ob}}}{\lambda_{\mathrm{em}}} - 1, \tag{1.4}$$

where the time of observation, t_{ob}, is taken to be today where $z = 0$.

The observed redshift distance relation is such that the change in the

wavelength of light emitted by distant galaxies, as characterised by their redshifts, is proportional to the distance to those galaxies. The relationship was first observed through the joint efforts of Hubble, Slipher and Humason and is alternatively expressed by the Hubble law, which states that the speed of recession of a galaxy, v, is linearly related to its proper distance, d, by

$$v = H_0 d, \tag{1.5}$$

where H_0 is the Hubble constant, the value of the Hubble parameter, H, today. This relation has been verified by numerous observational studies and the exact value of the Hubble constant has been continuously refined through successive observational surveys. The Hubble constant is commonly expressed by the dimensionless constant, h, where

$$H_0 = 100\, h\, \mathrm{km\, s^{-1}\, Mpc^{-1}}. \tag{1.6}$$

The full-mission Planck observations find $h = 0.678 \pm 0.009$ (Planck Collaboration et al. 2015). Due to the expansion of the Universe, the distance between us and a galaxy between two times t_1 and t_2 increases proportionally to the scale factor:

$$\frac{d(t_1)}{d(t_2)} = \frac{a(t_1)}{a(t_2)}. \tag{1.7}$$

Since the rate of change of the distance from us to a particular galaxy is equivalent to that galaxy's recession velocity, it can be shown that the Hubble parameter is given by

$$H(t) = \frac{\dot{a}(t)}{a(t)}. \tag{1.8}$$

Whilst the distance in the Hubble law refers to 'proper distance', it is useful to define a comoving distance, r_{co}, related to proper distance by

$$d(t) = a(t) r_{\mathrm{co}}. \tag{1.9}$$

In this way comoving distances normalise proper distances to the size they would have today, where today $a = 1$, and do not vary with time. Using the FLRW metric of Eq. 1.1, it can be shown that comoving distance is related to redshift

by

$$r_{\rm co}(z) = \int_0^z \frac{c}{H(z)} dz. \tag{1.10}$$

To understand the history of Universe it is necessary to understand the evolution of the scale factor as a function of time. Within general relativity, and assuming the matter and energy within the Universe can be modelled as a perfect fluid, the solutions of Einstein's equations provide solutions for $a(t)$, commonly referred to as Friedmann's equations. Perfect fluids are fluids which are isotropic in their rest frame, can be completely characterised by their pressure, p, and density, ρ, and frequently taken to satisfy

$$p = w\rho c^2, \tag{1.11}$$

where w is the dimensionless equation of state of the fluid. Friedmann's equations take the form

$$\frac{\dot{a}^2}{a^2} = \frac{8\pi G}{3}\rho - \frac{kc^2}{a^2} \tag{1.12}$$

$$\frac{\ddot{a}}{a} = -\frac{4\pi G}{3}\left(\rho + 3\frac{p}{c^2}\right), \tag{1.13}$$

where G is the gravitational constant. Taking the derivative of Eq. 1.12 with respect to time allows the two equations to be combined to provide an equation of energy conservation,

$$\dot{\rho} = -3H\left(\rho + \frac{p}{c^2}\right), \tag{1.14}$$

where the result $H \equiv \dot{a}/a$ has been used. Integration of Eq. 1.14 gives

$$\rho \propto a^{-3(1+w)}. \tag{1.15}$$

There are three constituents to consider, each obeying Eq. 1.15 with their own individual (constant) equation of state, w, whose combined densities make up the total energy density of the Universe: non-relativistic matter, with $w = 0$, radiation, with $w = \frac{1}{3}$, and the energy density of the vacuum itself, with $w = -1$. Assuming that the vacuum contributes to the total energy density of the Universe is equivalent to introducing a cosmological constant, Λ, to Einstein's equations and it can be seen that its equation of state $w = -1$ leads to a constant energy

density over time. The total energy density of the Universe then evolves as

$$\rho(a) = \rho_{\mathrm{m}} a^{-3} + \rho_{\mathrm{r}} a^{-4} + \rho_{\Lambda}, \tag{1.16}$$

where ρ_{m}, ρ_{r} and ρ_{Λ} are the current energy densities of matter, radiation and Λ respectively.

It is useful to define a critical density as

$$\rho_{\mathrm{crit}} = \frac{3H^2}{8\pi G}, \tag{1.17}$$

and introduce the density parameter, Ω, as the ratio between the current density and this critical density:

$$\Omega = \frac{\rho}{\rho_{\mathrm{crit}}} = \frac{8\pi G}{3H^2}\rho. \tag{1.18}$$

This critical density is chosen as it allows Eq. 1.12 to be written as

$$\Omega - 1 = \frac{kc^2}{H^2 a^2}. \tag{1.19}$$

Hence, determining whether the critical density parameter is greater than, equal to, or less than 1 fixes the sign of k, determining the curvature of the Universe. The assumption $k = 0$ required for a flat Universe therefore requires $\Omega = 1$.

The critical density parameter can again be split up into the 3 constituents,

$$\Omega = \Omega_{\mathrm{m}} + \Omega_{\mathrm{r}} + \Omega_{\Lambda}, \tag{1.20}$$

where Ω_{i} is the ratio between the energy density of the constituent, ρ_{i}, and the critical density. Combining Eq. 1.12, Eq. 1.16 and Eq. 1.4 allows the evolution of the Hubble parameter with redshift to be expressed as

$$H(z) = H_0[\Omega_{\mathrm{m0}}(1+z)^3 + \Omega_{\mathrm{r0}}(1+z)^4 + \Omega_{\Lambda 0}]^{\frac{1}{2}}, \tag{1.21}$$

when $k = 0$ and Ω_{i0} are the values of the critical density parameters today.

1.1.2 A brief history of the Universe

The theoretical framework laid out in the previous section indicates that the Universe will transition through three cosmic eras. The evolution of the scale

factor, which describes the expansion of the Universe, is driven by the evolution of the energy density. The energy density of matter and radiation decrease with time whilst the energy density of the vacuum remains constant. Hence the relative importance of the three types of energy density changes with time. The Universe was initially radiation dominated, then experienced a matter dominated era, and is now said to be Λ-dominated. In the limiting cases where the non-dominant contributions to the energy density are assumed negligible, the evolution of the scale factor with time is described in the three cosmic eras by:

- Radiation dominated: $\rho(a) \propto \rho_{\mathrm{r}} a^{-4} \rightarrow a(t) \propto t^{1/2}$
- Matter dominated: $\rho(a) \propto \rho_{\mathrm{m}} a^{-3} \rightarrow a(t) \propto t^{2/3}$
- Λ-dominated: $\rho = \rho_{\Lambda} \rightarrow a(t) \propto \exp(t)$.

It can be seen that all cases imply an expanding Universe, where a increases with time. Extrapolating this back to early times finds a singularity, which allows the approximate age of the Universe to be calculated as 13.8 billion years. This description of the dynamics of the Universe, starting from an initial singularity, is often called the "Big Bang" model. This model is widely accepted and is now strongly supported by observational evidence, most notably the cosmic microwave background (CMB), the abundance of light elements (Olive & Skillman 2004) and the aforementioned redshift-distance relation supporting the idea of an expanding Universe.

The CMB refers to the near-homogenous microwave radiation accidentally detected by Penzias and Wilson in 1965 (Penzias & Wilson 1965). The discovery remains one of the most significant developments in modern cosmology. In the Big Bang model it is understood to be the relic radiation left over from the transition between the radiation- and matter-dominated eras. At early times, the Universe was much hotter and denser than it is today and consisted of an ionised plasma of mostly electrons, protons and photons. The Universe was essentially opaque at these early times as the photons were constantly being scattered by the electrons and protons in the plasma. As the Universe expanded over time it cooled, until the temperature was low enough for the protons and electrons to combine, forming neutral atoms. This is often called the recombination epoch, and after this the Universe became transparent allowing photons to propagate freely through space. It is estimated that the recombination epoch occurred when the Universe was

approximately 380,000 years old, at a redshift of $z = 1100$ (Dodelson 2003). The CMB is the result of the photons that have been travelling towards us since the time at which radiation became decoupled from matter, with their wavelengths stretched into the microwave range due to the expansion of space.

1.1.3 Inflation, ΛCDM and the constituents of the Universe

Despite the significant evidence in support of the Big Bang model, there are a number of features of the Universe which are not well explained by this model alone. Firstly, the flatness problem: why is it that the density of the Universe is so close to the critical density required for a flat Universe which faces neither perpetual expansion nor an eventual collapse? Secondly, the horizon problem: it can be shown that there are causally disconnected regions of the Universe, and hence the global homogeneity observed is not explained within the Big Bang scenario. Finally, there is the question of initial conditions: the departure from small-scale homogeneity manifested by the collapsed objects seen in the Universe today requires initial perturbations which are not a part of the basic Big Bang model.

In 1981 Guth proposed the idea of inflation, initially as a solution to the monopole problem, and noted that inflation may also provide solutions to some of the above problems (Guth 1981). In the inflationary scenario, the Universe is said to have undergone a short period of quasi-exponential expansion, after which it continued to expand, as predicted by the Big Bang model, but at a much slower rate. This initial inflationary period allows the entire observable Universe to have been in causal contact and provides an explanation for the large-scale homogeneity and isotropy observed today. The rapid expansion of inflation also solves the flatness problem as it is a general result of inflationary theories that as a increases enormously, Ω is driven towards unity. Inflation can also provide the initial perturbations necessary for the structures that we see today to have formed. Quantum uncertainties in the density field would have been stretched to macroscopic scales allowing them to 'freeze in' (Lyth & Liddle 2009). Although inflation provides solutions to many of these problems, the exact mechanisms involved are not clear and inflationary theories remain an active field of study.

The lambda cold dark matter (ΛCDM) model is the currently favoured

standard model of cosmology which encompasses the Big Bang model and general relativity and allows for inflation. As mentioned in Section 1.1.1, three components contribute to the total energy density of the Universe. Due to the expansion of space the energy density of radiation today is extremely low. The bulk of the energy within our Universe is then thought to be in the form of matter or dark energy, as represented by the cosmological constant, Λ.

There is reason to believe that matter can be of two types: baryonic matter and dark matter. Whilst baryonic matter encompasses all the matter we see, dark matter interacts only through gravity. Its existence was first postulated by Zwicky in the 1930s to explain mass discrepancies seen in the Coma cluster. Later, the flattened rotation curves of many spiral galaxies provided further evidence in support of dark matter (van Albada et al. 1985). Since then, observations of the CMB, and the peaks in the CMB power spectrum, have provided strong evidence of dark matter. Today many further observational probes, including gravitational lensing, have contributed to the large body of evidence in favour of dark matter. However, although many experiments are attempting to do so, dark matter has not yet been directly detected and its exact nature is not fully understood. It is currently believed the Universe is composed of $\sim 5\%$ baryonic matter, $\sim 27\%$ dark matter, and that the remaining $\sim 68\%$ is dark energy (Ade et al. 2014). The proposed existence of dark energy is the result of adding the cosmological constant, Λ, to the FLRW metric in order to explain the observed expansion of space. It is even less well understood than dark matter and remains an active field of study within cosmology.

1.2 Structure formation

Section 1.1 laid out the theoretical framework of the prevailing model for our homogenous, smooth and expanding Universe. However, whilst it is common to postulate that the Universe is homogeneous on large scales, the very existence of our solar system and of the many galaxies that have been observed show that the Universe is far from homogenous on small scales. The field of structure formation attempts to understand this departure from homogeneity by modelling the gravitational collapse of matter.

For some time there was an uncertainty over whether structure formed

via a 'top-down' approach or by 'bottom-up' formation, and there were many competing models of structure formation (Peebles & Silk 1990). The top-down approach postulated that large-scale structures, the size of galaxy clusters, formed first, and later fragmented into smaller structures such as galaxies. It was suggested that small-scale fluctuations were smoothed out by the motions of particles, and hence this model was mostly associated with a hot dark matter dominated Universe. However in the past few decades the consensus has moved towards the cold dark matter standard model, which supports the bottom-up formation of structure. Hence it is commonly assumed that structure forms via a hierarchical process, whereby small-scale fluctuations collapse first and later merge to produce large structures. The observational evidence of recent years now strongly supports this model of structure formation (Steinmetz & Navarro 2002).

Since collapse is well outside of the linear regime due to its large overdensities, it is necessary to make some simplifying assumptions. The most common simplification is to consider a spherically symmetric overdensity. An overdensity refers to a density contrast, δ, which is greater than 0. The density contrast is defined throughout this work as

$$\delta = \frac{\rho - \bar{\rho}}{\bar{\rho}}, \tag{1.22}$$

where ρ is the local matter density and $\bar{\rho}$ is the average matter density of the Universe. It can then be said that we are in the linear regime when $\delta \ll 1$. We are also required to postulate a set of initial conditions, from which gravitational collapse began to proceed. The standard picture, which stems from Guth's theory of inflation (Guth 1981), is that the Universe initially had a very smooth density field, with $\delta \ll 1$ everywhere, and with small primordial density fluctuations which originated from a Gaussian random process.

Due to the Gaussian nature of these primordial fluctuations, it is often beneficial to work in Fourier space, with the Fourier counterpart of the density contrast, δ_k, defined as

$$\delta_k = \frac{1}{V} \int \delta(t, \mathbf{x}) e^{-i\mathbf{k}.\mathbf{x}} d^3x, \tag{1.23}$$

with V representing the volume element being integrated over. The power

spectrum, $P(k)$, is a useful way to characterise such a field of density fluctuations, and it is essentially the ensemble average of these Fourier components:

$$P(k) = |\delta_k|^2. \tag{1.24}$$

The primordial fluctuations in their most simple form produce a scale-invariant spectrum, the Harrison-Zel'dovich spectrum (Harrison 1970, Zeldovich 1972) given by

$$P(k) = Ak, \tag{1.25}$$

though it should be noted that this simple model does not hold within inflationary scenarios which introduce a characteristic scale. At subsequent times, the shape of the small-scale end of the power spectrum will be highly dependent on the constituents and dynamics of the expanding Universe. This is often encoded in a transfer function, $T(k)$, and the form of the power spectrum today allows us to learn about the physics at work between the inflationary era and later times.

1.2.1 Spherical collapse

The problem of a spherically symmetric overdensity collapsing in a matter-dominated Universe is relatively simple to solve. It is equivalent to considering the behaviour of a closed sub-universe. To do this, Friedmann's equation (Eq. 1.12) describing the expansion of space in a homogenous and isotropic universe is used. It is easiest to proceed after making a transformation to conformal time, η, by

$$d\eta = c\frac{dt}{a}. \tag{1.26}$$

The Friedmann equation then becomes

$$\left(\frac{\mathrm{d}a}{\mathrm{d}\eta}\right)^2 = \frac{8\pi G\rho a^4}{3c^2} - ka^2. \tag{1.27}$$

By introducing a parameter, $a_\star$, (which is constant in a matter dominated universe) defined as

$$a_\star = \frac{4\pi G\rho a^3}{3c^2} = \frac{H^2\Omega a^3}{2c^2}, \tag{1.28}$$

the Friedmann equation can be rewritten as:

$$\left(\frac{\mathrm{d}a}{\mathrm{d}\eta}\right)^2 = 2a_\star a - ka^2. \tag{1.29}$$

When $k = 1$, it can be shown that a solution is given by

$$a = a_\star(1 - \cos\eta), \tag{1.30}$$

and integrating over conformal time to return to cosmic time, t, it can be shown that, to within an arbitrary constant,

$$ct = a_\star(\eta - \sin\eta). \tag{1.31}$$

If we assumed $k = 0$, we would have reached the simpler $a \propto t^{2/3}$ solution, and for $k = -1$, the solutions are $a = -a_\star(1-\cosh\eta)$ and $ct = -a_\star(\eta-\sinh\eta)$. These

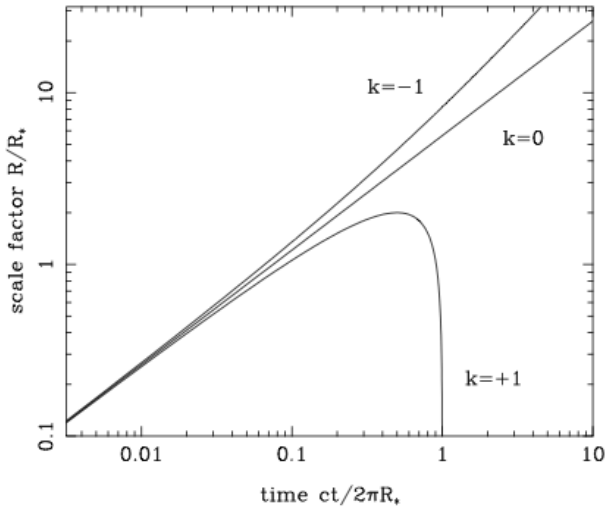


Figure 1.1 The evolution of the scale factor (here $R \equiv a$ and $R_\star \equiv a_\star$) as a function of time for open (k= −1), critical (k= 0) and closed (k= 1) matter-dominated cosmological models (Peacock 2001).

solutions are plotted in Fig. 1.1.

It is the $k = +1$ solution which most closely resembles the behaviour of collapsing matter, since the $-kc^2$ part of the Friedmann equation represents negative total energy and hence a bound universe which must collapse back in

on itself. In reality, our Universe does not correspond to any of these solutions since we have neglected any vacuum energy. Carroll et al. (1992) present a more thorough discussion of the solutions in realistic cosmological models.

For the application of the above analysis to spherical collapse, we need the more general cycloid solutions for the $k = +1$ case:

$$a(\theta) = A[1 - \cos(\theta)] \qquad t(\theta) = B[\theta - \sin(\theta)]. \tag{1.32}$$

Here conformal time has been replaced by a parameter, θ, often called the development angle, and A and B are constants. From $\ddot{r} = -GM/r^2$, it can be shown that $A^3 = GMB^2$. Expansion of the trigonometric terms allows us to approximate the form of $a/t^{2/3}$, which, by substitution of the approximate form of $t(\theta)$, leads us to

$$a(t) \approx \frac{A}{2}\left(\frac{6t}{B}\right)^{2/3}\left(1 - \frac{1}{20}\left(\frac{6t}{B}\right)^{2/3}\right). \tag{1.33}$$

By substituting $\rho = 3M/4\pi a^3$ and $\bar{\rho} = 3H^2/8\pi G$ directly into the definition of δ given in Eq. 1.22, and by using the matter dominated form of the Hubble parameter, $H = 2/3t$, we can derive an equivalent expression for $a/t^{2/3}$:

$$\frac{a}{t^{2/3}} \approx \left(\frac{9GM}{2}\right)^{1/3}\left(1 - \frac{\delta}{3}\right). \tag{1.34}$$

Comparison of Eq. 1.33 and Eq. 1.34 allows us to identify the density perturbation within the sphere:

$$\delta \approx \frac{3}{20}\left(\frac{6t}{B}\right)^{2/3}. \tag{1.35}$$

From the trigonometric form of $a(\theta)$ we can identify two important epochs, the moment of turnaround and the moment of collapse. The maximum radius and hence the point of turnaround occurs when $\theta = \pi$. We can derive the density contrast at turnaround by substituting in our expressions for ρ and $\bar{\rho}$, and letting A^3/B^2 replace GM. From here the expression for $\rho/\bar{\rho}$ simplifies to a

more manageable form:

$$\frac{\rho}{\bar{\rho}} = \frac{9A^3t^2}{2B^2a^3} = \frac{9(\theta - \sin\theta)^2}{2(1-\cos\theta)^3}. \tag{1.36}$$

Thus the density contrast at turnaround is $1 + \delta = 9\pi^2/16 = 5.55$, and hence the fluctuation has already entered the non-linear regime before it begins to collapse. Extrapolation of the linear theory result shown in Eq. 1.35 gives us a value of δ at turnaround of 1.06. To quantify the density contrast at which the object has fully collapsed is not as straightforward; in this approximation of a perfectly spherical perturbation the collapse would proceed until an infinite density is reached when $\theta = 2\pi$. However one can argue that in reality small deviations from perfect symmetry would cause the object to virialize before it reached a singularity, when dissipative physics converted the kinetic energy of collapse into random motions. Since virialization would not occur instantly, it is common to take the time of virialization, and hence the time of collapse, to be equivalent to the time at which a perfect sphere would reach a singularity, when $\theta = 2\pi$. This suggests that the density contrast at collapse is 177, and the linear theory value, δ_c, is equal to 1.686. Importantly, it implies that all virialized objects will have the same density contrast regardless of their mass.

The same result can be reached by energy conservation arguments starting from $\mathrm{U}_{\mathrm{R_{max}}} = \mathrm{U_{vir}} + \mathrm{T_{vir}}$, since $\mathrm{T}_{\mathrm{R_{max}}} = 0$. Where U and T are the potential and kinetic energies respectively, and the subscripts $_{\mathrm{R_{max}}}$ and $_{\mathrm{vir}}$ refer to the time of turnaround and time of virialization respectively. Since $\mathrm{U} \propto \mathrm{R}^{-1}$, the virial theorem tells us that $2\mathrm{T_{vir}} = -\mathrm{U_{vir}}$ and hence $\mathrm{U}_{\mathrm{R_{max}}} = \mathrm{U_{vir}}/2$ and the radius at virialization is half the radius at turnaround. If we again take the time of collapse (time of virialization) to be when $\theta = 2\pi$, we find the time at collapse is twice the time at turnaround, and then, for a matter dominated universe where $a \propto t^{2/3}$, we can say that the average density of the Universe has decreased by a factor of 4 between turnaround and virialization. Combining this with the decrease in radius, which increases the density by a factor of $2^3 = 8$, we find $\rho_{\mathrm{vir}}/\bar{\rho} = 4 \times 8 \times (1 + \delta_{\mathrm{R_{max}}}) = 32 \times 5.55 = 178$ and again we see that the density contrast of the collapsed perturbation, δ_{vir}, is equal to 177.

1.2.2 Press Schechter formalism

One of the most fundamental characteristics of a bottom-up structure formation scenario is the mass function – the number density of objects as a function of their mass, M. In 1974, Press and Schechter developed a theoretical model for the mass function (Press & Schechter 1974, hereafter PS). Despite its relative simplicity and some arguably incorrect assumptions, the PS formalism has been shown to agree well with the output of numerical simulations over some mass ranges (Lacey & Cole 1994), but has been criticised for under-predicting the abundance of high mass halos and over-predicting that of low mass halos (Sheth & Tormen 1999). Various extensions to the theory have been discussed in the literature (see e.g. Peacock & Heavens 1990), the Sheth-Tormen (Sheth & Tormen 1999) mass function in particular has been shown to provide better fits (Reed et al. 2007).

The PS theory in essence describes the mass function of virialized objects resulting from dissipationless collapse of initially cold matter. It draws on the assumptions discussed above, assuming the initial density fluctuations were Gaussian and utilising the density contrast derived in section 1.2.1, δ_c, which describes the threshold density contrast for a spherically symmetric perturbation to virialize. Its simplicity stems from the further assumption that the amplitude of the large-wavelength modes of the otherwise non-linear density field will remain close to the linear-regime, allowing PS to develop their accurate, yet linear, theory.

With these assumptions, PS argued that the fraction of mass within collapsed objects of mass M or greater can be derived from the fraction of the density field with a density contrast greater than the linear theory threshold for collapse when it is smoothed with a Gaussian filter of various scales, such that the smoothed density field has variance $\sigma(R_s)$. The smoothing scale, R_s, provides a way to estimate the size and hence the mass of the collapsed object by $M \sim \rho R_s^3$. PS postulated that the probability that a given point lies in a region with $\delta > \delta_c$ is given by:

$$P(\delta > \delta_c | R_s) = \frac{1}{\sqrt{2\pi\sigma^2}} \int_{\delta_c}^{\infty} \exp\left(-\frac{\delta^2}{2\sigma^2}\right) d\delta \equiv \frac{1}{2}\left[1 - \mathrm{erf}\left(\frac{\delta_c}{\sqrt{2}\sigma}\right)\right]. \quad (1.37)$$

Introducing the parameter $\nu = \delta_c/\sigma$, it can be shown that the fraction, $F(> M)$,

of the Universe in collapsed objects with mass greater than $M(R_s)$ is given by

$$F(>M) = \frac{1}{2}\left(1 - \mathrm{erf}\left(\frac{\nu}{\sqrt{2}}\right)\right) \equiv \frac{1}{2}\mathrm{erfc}\left(\frac{\nu}{\sqrt{2}}\right). \tag{1.38}$$

However it is noticeable from Eq. 1.38 that even as $\sigma \to \infty$ and $\nu \to 0$, the collapse fraction is still only 1/2, whereas we expect that every particle should end up as part of some halo with $M > 0$. For this reason PS multiplied the resulting expression by 2, without having a theoretical justification at the time. Extensions to the PS formalism which account for this missing factor of 2 have since been developed, as discussed below.

The resulting formula allows us to express the PS multiplicity function, $M^2 f(M)/\rho_0$, in terms of the mass function, $f(M)$, where $f(M)dM$ is the comoving number density of objects in the mass range dM. The multiplicity function expresses the fraction of mass in objects within a unit range of $\ln(M)$. We can express the mass function in terms of the above $F(>M)$ by noting that

$$\frac{M}{\rho_0} f(M) = \left|\frac{dF(>M)}{dM}\right| \tag{1.39}$$

and hence

$$\frac{M^2 f(M)}{\rho_0} = \left|\frac{dF(>M)}{d\ln M}\right| = \left|\frac{d\ln\sigma}{d\ln M}\right| \sqrt{\frac{2}{\pi}}\nu \exp\left(-\frac{\nu^2}{2}\right). \tag{1.40}$$

1.2.3 Excursion set formalism

The excursion set formalism (Bond et al. 1991, Bower 1991), often referred to as the extended Press-Schechter formalism, extends the PS approach to structure formation by considering the correlations between different mass scales. In standard PS theory, if the density contrast on a particular smoothing scale is below the threshold it is not considered to be collapsed, yet it may be above the threshold on some larger smoothing scale. It is necessary, then, to find the largest smoothing scale which causes the density contrast to exceed the threshold for collapse.

Since the development of the PS approach it has been argued that if we consider the smoothing in k-space, it is equivalent to a 'random-walk' in δ, and hence for each trajectory that lies above the threshold at R_s, there is an equivalent

trajectory which is its mirror image around δ_c. Bond et al. (1991) argued that regions which are not above the threshold on some smoothing scale may lie above the threshold on a larger smoothing scale and should also be counted. Hence, whilst PS considers the fraction of the density field above the threshold for a given smoothing scale, the excursion set formalism calculates the fraction of the trajectories with a first upcrossing of the threshold for the given smoothing scale. Furthermore, since each trajectory must eventually cross the threshold for some arbitrarily large smoothing scale, the excursion set formalism predicts that every particle is in a halo of some arbitrarily low mass. It can be shown that the mass function resulting from this alternative approach is identical to that of the PS formalism, without the need for the unjustified factor of 2.

1.2.4 Bias and the peak background split

It has been recognised for many years that the distribution of galaxies need not follow exactly the distribution of the dark matter. A mechanism for this bias was first proposed by White et al. (1987), where it was noted that the probability of a collapsed object of mass M forming in a certain location is affected by the large-scale density field, which is to be expected since Section 1.2.1 showed us that the probability of collapse depends on the initial density field. Similarly, it is expected that dark matter halos are also biased tracers of the underlying dark matter field, and a theoretical model for this halo biasing within the PS formalism has been developed (Cole & Kaiser 1989, Mo et al. 1996). It has been argued that the large scale density component, ϵ, adds to the small scale component, δ, effectively reducing the threshold which needs to be reached by δ in order to bring about collapse. That is, $\delta_c \rightarrow \delta_c - \epsilon$. This changes the number density predicted by the PS formalism in a region of large-scale density contrast ϵ by

$$n \rightarrow n + \frac{dn}{d\nu}\frac{d\nu}{d\epsilon}\epsilon = n\left[1 + \epsilon\frac{(\nu^2 - 1)}{\sigma\nu}\right]. \tag{1.41}$$

From Eq. 1.41 we can identify the Lagrangian bias, b_L, in the number density of halos, where $\delta(n)/n = b_L\epsilon$.

$$b_L = \frac{(\nu^2 - 1)}{\nu\sigma}. \tag{1.42}$$

To estimate the Eulerian bias, the dynamical effects must be taken into account; the volume will be reduced by a factor $(1-\epsilon)$ in order to contain the same mass. Therefore, assuming $\epsilon \ll 1$, the Eulerian space bias, b, is given by

$$b = 1 + \frac{(\nu^2 - 1)}{\nu\sigma}. \tag{1.43}$$

The derivation of the Eulerian space bias is discussed in more detail by Sheth and Tormen, who, in a later paper, also provide an improved expression for the bias derived from the Sheth-Tormen mass function (Sheth & Tormen 1999, Sheth et al. 2001):

$$b = 1 + \frac{a\nu^2 - 1}{\delta_c} + \frac{2p}{\delta_c[1 + (a\nu^2)^p]}, \tag{1.44}$$

where $a \sim 0.75$ and $p \sim 0.3$ in order to match numerical simulations.

1.2.5 The Zel'dovich approximation

All theories presented so far have been based within the linear-regime – on small scales, in the non-linear regime, the complexities of gravitational instability make exact analytical analysis near impossible. The remaining options are to study numerical simulations of the non-linear regime, or to make approximations which allow us to find solutions for the general case. One such approximation was proposed by Zel'dovich (1970), in which an approximate kinematic solution was given for the growth of perturbations during the expansion of matter without pressure, when the initial matter distribution is considered to be homogenous and collisionless.

The approach relies on the approximation that once the initial displacement of particles is known they continue to move in this direction indefinitely, or at least until particle trajectories cross, after which the solution is no longer valid. It provides a relation between Lagrangian (initial) co-ordinates, $\mathbf{q}$, the Eulerian proper coordinates, $\mathbf{r}$, and the comoving coordinates, $\mathbf{x}$, of the particles at a time t:

$$\mathbf{r}(\mathbf{q}, t) = a(t)\mathbf{x}(\mathbf{q}, t) = a(t)[\mathbf{q} + D(t)\mathbf{f}(\mathbf{q})]. \tag{1.45}$$

Here $a(t)$ is the usual cosmological scale factor, $D(t)$ the linear density growth factor and the initial displacement field is denoted by $\mathbf{f}(\mathbf{q})$, which is a function of the initial Lagrangian coordinate only (Padmanabhan 2002).

Hence, the Zel'dovich approximation provides a way to extrapolate from a given initial displacement field to the approximate positions of the particles today. Furthermore, the conservation of mass principle can be used to deduce the approximated perturbed density field today. Mass conservation implies

$$\rho(\mathbf{r}, t)d^3\mathbf{r} = \overline{\rho}d^3\mathbf{q}, \tag{1.46}$$

therefore, by using the Jacobian to compare volume elements, the density at time t is given by

$$\rho(\mathbf{r}, t) = \overline{\rho}\left[\det\left(\frac{\partial r_i}{\partial q_j}\right)\right]^{-1} = \frac{\overline{\rho}/a^3}{\det(\partial x_i/\partial q_j)} = \frac{\overline{\rho}}{\det\left[\delta_{ij} + D(t)(\partial f_i/\partial q_j)\right]}. \tag{1.47}$$

The Jacobian in the denominator of Eq. 1.47 is the determinant of a real symmetric matrix which can be diagonalised to yield a set of eigenvalues and principal axes as functions of $\mathbf{q}$, or, by use of Eq. 1.45, as functions of $\mathbf{x}$. If these eigenvalues are found to be α, β, γ, and the off diagonal elements of the matrix in the denominator, $D(t)(\partial f_i/\partial q_j)$, are assumed to negligible compared to the diagonal $(1 + D(t)(\partial f_i/\partial q_i))$ elements, the perturbed density field in comoving coordinates is given by

$$\rho(\mathbf{x}, t) = \overline{\rho}\left[(1 - D(t)\alpha(\mathbf{x}, t))\,(1 - D(t)\beta(\mathbf{x}, t))\,(1 - D(t)\gamma(\mathbf{x}, t))\right]^{-1}. \tag{1.48}$$

Eq. 1.48 shows that a positive eigenvalue (α, β, or γ greater than zero) leads to collapse along the corresponding eigenvector whereas a negative eigenvalue leads to expansion. Although the ratios of these eigenvalues are constant, the ratios of $(1 - D(t)\alpha)/(1 - D(t)\beta)$, for example, are not constant. This causes the collapse to preferentially proceed along the direction of the strongest gravitational potential, and hence collapse becomes progressively lower dimensional with the passage of time (Yoshisato et al. 2006).

The limitations of the Zel'dovich approximation are clear from Eq. 1.48; once an eigenvalue approaches 1, the density in that dimension diverges, and if all eigenvalues reach 1 the theory predicts an infinite density. This stems from the assumption that particles continue to move along their initial trajectory; if a

particle reaches a maximum in the density field, this assumption would have it continue past the overdensity – it is at this point of shell crossing at which the theory breaks down. The Zel'dovich approximation is thus valid up until the quasi-linear regime, for eigenvalues ≤ 1.

Despite its simplicity, the Zel'dovich approximation is surprisingly accurate (see e.g. Yoshisato et al. 2006 for comparisons of the approximation to exact numerical solutions) and is in fact an exact solution in one dimension. Accordingly, it is widely used throughout cosmology. For example, it is used in the construction of initial conditions for N-body simulations (Bertschinger 1998), and conversely, in the derivation of estimated initial conditions given the observed distribution of galaxies in the local Universe (e.g. Nusser & Dekel 1992, Dekel 1994). It has been argued that the success of the Zel'dovich approximation is due to the fact that the density contrast is a nonlinear function of displacement; the displacement may be small and within the linear regime, yet the density contrast may be large and nonlinear. This allows the Zel'dovich approximation to successfully extend beyond the linear regime at which point most analytical models fail. Despite its success, other authors have proposed extensions to the Zel'dovich approximation, which is essentially first-order Lagrangian perturbation theory, to try to improve its accuracy or range of validity (e.g. Bernardeau et al. 2002, Brainerd et al. 1993, Melott et al. 1994).

1.3 The cosmic web

The so called cosmic web refers to the visually striking way in which the cosmic density field separates into distinct structures with different dimensionality. Whilst it was postulated above that the Universe is homogeneous on large scales, on megaparsec scales the matter distribution is far from uniform. In fact, it forms a complex and intricate web-like pattern. The first hint of the existence of such a web-like pattern came with early redshift surveys providing a view of the local Universe (Gregory & Thompson 1978, de Lapparent et al. 1986). Since then, modern surveys such as 2dFGRS (Colless et al. 2003a) and SDSS (Abazajian et al. 2003) have measured the locations of hundreds of thousands of galaxies and confirmed the existence of web-like structures – the largest structures in the known Universe. In addition to the local Universe, these patterns have been seen

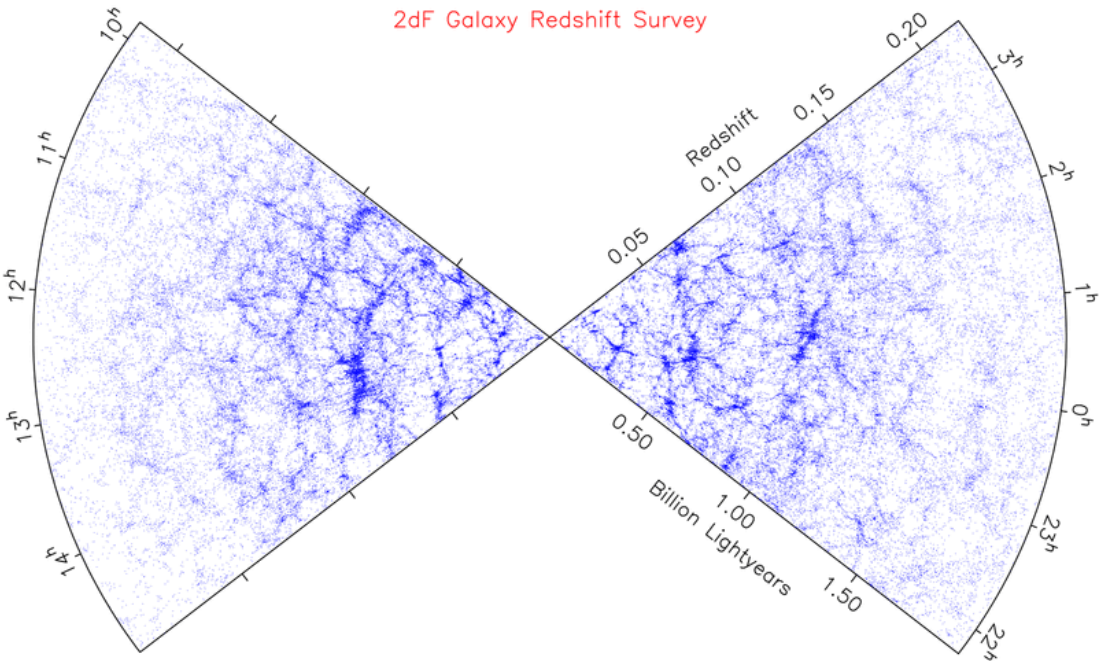


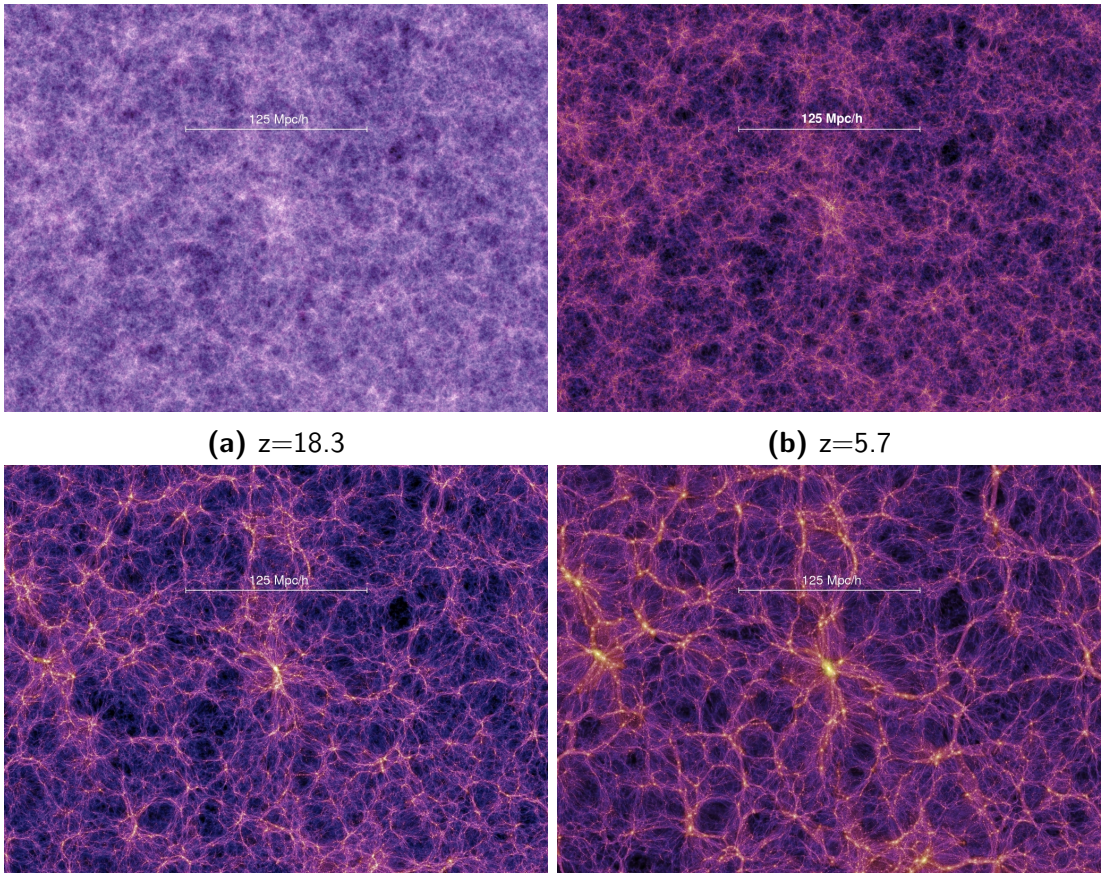
Figure 1.2 The galaxy distribution measured by the 2dF redshift survey, illustrating the web-like nature of the Universe. (from Colless et al. 2003b)

to persist into the high-redshift Universe. The Subaru survey has found evidence of filamentary features at $z \sim 3.1$, indicating the existence of cosmic structure at early cosmic epochs (Hayashino et al. 2004).

The galaxy distribution measured by these surveys allows maps of the Universe, such as the 2dF slice shown in Fig. 1.2, to be created. It can be seen from Fig. 1.2 that the galaxy distribution exhibits filamentary structure – 1-dimensional lines of galaxies. There are dense and compact regions at the intersection of the filaments, where many galaxies reside. These maps also exhibit vast, almost empty regions with very few galaxies, called voids. When viewing the distribution in 3-dimensions, it is also possible to distinguish walls – flat, 2-dimensional sheets of galaxies. For example, a vast planar arrangement of galaxies dubbed the Great Wall has been identified in the local Universe (Geller & Huchra 1989). Since then, even larger planar structures have been found at higher redshift, for example the Sloan Great Wall identified by SDSS. These constitute the four components of the web that are most often considered: 3D voids and knots at either extreme of overdensity, connected by 2D sheets and 1D filaments.

In addition to the observational evidence of the cosmic web, numerous large-scale numerical simulations have seen the web-like nature of the matter distribution reproduced as a natural consequence of standard ΛCDM cosmology (Shandarin 2011, Einasto et al. 2011, Sousbie et al. 2011, Aragon-Calvo 2012). For example, the Millennium simulation traced the evolution of more than 10 billion particles from early cosmic epochs to the present day (Springel et al. 2005). Fig. 1.3 shows 15 h^{-1}Mpc thick slices of the density field from the Millennium simulation for four redshift snapshots from $z = 18.3$ to today. The evolution of the density field shows a hierarchical build up of the cosmic web over cosmic time as the components take sharper and more distinct forms. The density contrasts amplify over cosmic time producing distinctly underdense regions and bright dense clusters of mass intersected by wispy filamentary structures.

Structure formation is understood to occur as a result of the gravitational collapse of matter (Zel'dovich 1970, Peebles 1980). Initial anisotropies in the density field are amplified by gravitational forces giving rise to structures such as the cosmic web. Overdense regions will exert stronger gravitational forces on their surroundings and accrete matter from their surroundings accordingly, amplifying



(a) z=18.3 **(b)** z=5.7

(c) z=1.4 **(d)** z=0

Figure 1.3 Four slices of the density field from the Millennium ΛCDM dark matter simulation. The slices are 15 $h^{-1}\mathrm{Mpc}$ thick and show four redshift snapshots as indicated below each panel. The evolution of the density field reveals the hierarchical buildup of the cosmic web into distinct geometric features. Images from the V. Springel & Virgo consortium (Springel et al. 2005).

their overdensities and fuelling the continuation of the process. Contrastingly, underdense regions exert weaker gravitational forces than their surroundings and the average movement of matter will be away from the most underdense regions.

The hierarchical build up of the cosmic web can be understood within the Zel'dovich framework discussed in Section 1.2.5. Unless an overdensity is spherically symmetric, the Zel'dovich approximation showed how gravitational collapse accentuates anisotropy (Arnold et al. 1982, Hidding et al. 2014); collapse first proceeds along the direction experiencing the strongest gravitational force (the eigenvector corresponding to the largest eigenvalue of Eq. 1.48). This produces a 'Zel'dovich pancake', where structure has collapsed along one dimension leaving a 2D sheet. This is followed by collapse along the second strongest direction, producing the 1D structures seen as filaments. Collapse along the final dimension produces the nodes of the web – clusters or knots of extreme overdensity.

The components of the cosmic web represent the largest geometric structures in the Universe and reveal complex patterns of a highly organised nature. As they exist on scales of the order of megaparsecs, they represent the mildly non-linear regime in which the Universe is still in a state of moderate dynamical evolution. They can thus provide invaluable information on the matter distribution in the primordial Universe and on the process of cosmic structure formation. This realisation, along with the advent of large, detailed simulated and observational datasets in recent years, has led to the cosmic web becoming an active and fruitful field of study. Numerous methods of classifying the cosmic web have been proposed, some of which are discussed in more detail below.

1.3.1 Extracting the components of the web

Whilst the cosmic web is easy to distinguish by eye, to quantitatively define it is nontrivial due to the range of scales, shapes and dimensionality involved. Furthermore, there is no strict definition as to what constitutes the individual components of the web and its development is a continuous process rather than a discrete well-defined one. As with most scientific studies, a reproducible quantitative method of classification is required.

Early statistical tools aimed at describing the cosmic network have proved useful in comparing observations to theory. These include the two- and three-

point correlation functions (Peebles & Groth 1975, Peebles 1980) and Minkowski functionals (Mecke et al. 1994, Aragon-Calvo et al. 2010). Other attempts focus on extracting the 'skeleton' of the cosmic web in order to trace the filamentary pattern of the density field (Novikov et al. 2006, Stoica et al. 2010). The skeleton is determined by segments parallel to the gradient of the field, connecting saddle points to local maxima, and is based on Morse theory (Colombi et al. 2000, Eriksen et al. 2004). This skeleton approach has been applied to both numerical simulations and observational datasets in order to identify filaments (Sousbie et al. 2008, Sousbie et al. 2008b).

There have also been more empirical methods proposed to extract the skeletal nature of the galaxy distribution. For example minimal spanning tree methods (Barrow et al. 1985) based on graph theory find the shortest path connecting 'nodes' (usually galaxies or galaxy groups). These methods have proven to be useful tools for identifying filaments within observational surveys (Doroshkevich et al. 2004, Alpaslan et al. 2014), however they are closely connected to the nature of the data in a given survey.

In addition to these techniques focused on finding filaments, the Watershed Void Finder (Platen et al. 2007), based on the watershed algorithm, a technique for the segmentation of images, has been developed and shown to identify voids and their surrounding boundaries. Many other void finding algorithms exist and have been shown to be in good agreement with each other despite varying methods (see Colberg et al. (2008) for a comparison of void-finders).

Whilst there are many algorithms aimed at identifying particular components, or the skeleton of the web, alternative approaches aim to provide a classification analysis of space, where the geometric nature of every position in the full volume is classified. These most often characterise the cosmic web based on the gravitational field Hessian or the velocity shear field. The Hessian approach, strongly linked to the Zel'dovich approximation, was first proposed by Hahn et al. (2007). In this prescription, the eigenvalues of the Hessian (the tidal tensor) are used to indicate the dimensionality of collapse. This approach has seen the cosmic web environments classified in both simulated and observational datasets by many authors (Forero-Romero et al. 2009, Zhang et al. 2013, Alonso et al. 2015, Eardley et al. 2015). This method was recently extended by Leclercq et al. (2015) who used Bayesian decision theory to take the strength of data constraints into

account. Related approaches allow for the identification of structure on multiple scales, such as the Multi-scale Morphology Filter (Aragón-Calvo et al. 2007) and NEXUS+ (Cautun et al. 2013).

Hoffman et al. (2012) argued that the Hessian approach is not able to resolve the cosmic web at fine scales, as shown by high resolution N-body simulations. This motivated their use of the velocity shear tensor instead of the tidal tensor; whilst the two are essentially identical in the linear regime, they diverge in the fully non-linear regime where the velocity shear is better able to identify the fine structures of the web. Many authors have used the velocity shear tensor to classify the cosmic web in numerical simulations (Libeskind et al. 2013, Tempel et al. 2014, Metuki et al. 2015). However, the requirement of velocity information limits the applicability of this approach to observational datasets (see Lee et al. (2014) for an analysis of the velocity shear web at the position of the Virgo cluster). Recently some authors have utilised both the velocity and density information, using the full 6D phase space to probe the cosmic web (Shandarin et al. 2012, Abel et al. 2012, Falck et al. 2012). This allows for a more robust characterisation of the web but again it limits the applicability to observational surveys.

1.4 Cosmological datasets

1.4.1 Galaxy redshift surveys

In order to develop, test and refine all of the theoretical models presented thus far, it is necessary to compare the predictions of these models to the statistics of the observable Universe. Galaxy redshift surveys, along with other observational probes such as gravitational lensing and CMB surveys, play an important role in these comparisons. Over the past few decades, the advent of large, international galaxy redshift surveys has transformed the field of modern observational cosmology by providing detailed and expansive pictures of the Universe. Together with the ever larger numerical simulations, these surveys have been instrumental in shaping our understanding of large scale structure and in forming a primary consensus within the scientific community of the standard cosmological model – ΛCDM.

By measuring the change in wavelength of a galaxy's spectrum between emission and observation, Hubble's law allows for recessional velocities, and hence

redshifts, to be converted into radial distances. This, together with the angle on the sky, characterised by right-ascension (ra) and declination (dec) within the equatorial coordinate system, allows galaxy redshift surveys to provide us with a 3-dimensional map of structure within the Universe.

Notable recent surveys include the Sloan Digital Sky Survey[1] (SDSS), 2dF[2], and DEEP2[3]. Significant work and expertise is required in order to extract information from these large datasets; even once redshifts have been obtained from the many thousands of spectra, issues such as completeness and contamination must be considered before an accurate and reliable galaxy catalogue is achieved. It is then often beneficial to studies of large scale structure to compile a group catalogue, where galaxies are defined as being within a group due to their spatial proximity. The prevalent method of grouping nearby galaxies is the Friends-of-Friends technique (FoF) (Geller & Huchra 1983). This method defines a maximum distance, or linking length, r_{12}, between a member of a group and its closest neighbour within the same group (often set to be $r_{12} \approx 0.2n$, where n is the average distance between neighbouring galaxies). Groups are then classified as being those collections of galaxies with more than a specified number of members which meet the linking length criterium. These group catalogues are useful as they are often considered to be the visible, measurable counterparts of dark matter halos.

Galaxy redshift surveys primarily utilise either photometric or spectroscopic detection techniques. Photometric redshifts are obtained using broad-band photometry and multiple filter images; an estimate of the redshift is obtained by observing broad features such as the Balmer or Lyman continuum breaks or the overall shape of the spectrum. The main advantage of this method is that, because photometric surveys use significantly larger wavelength bins than spectroscopic surveys, a far shorter exposure time is needed to allow for the signal to noise ratio to reach sufficiently high levels. This means that photometric surveys are able to cover a wider and deeper field in much less time than spectroscopic surveys. On the other hand, spectroscopic surveys observe the entire spectrum of individual objects and split the light into narrow wavelength bins. This allows for narrow features such as emission or absorption lines in the spectra to be identified and

[1] `http://www.sdss.org/`
[2] `http://www.mso.anu.edu.au/2dFGRS/`
[3] `http://deep.berkeley.edu/`

thus the shift from the known rest wavelengths can be measured and used to obtain precise redshifts. The detection technique used in each redshift survey depends on the scientific aims of the project – while spectroscopic surveys provide accuracy incomparably greater than photometric surveys, this quality comes at the expense of quantity.

In addition to measuring the 3-dimensional distribution of galaxies, observational surveys can provide a wealth of further information from analysis of the galactic spectra. An estimate of the total amount of energy emitted by the galaxy provides a luminosity measurement. A related quantity, magnitude, characterising a galaxy's brightness on a logarithmic scale, can also be derived. The commonly used AB magnitude, m_{AB}, is defined as

$$m_{\mathrm{AB}} = -2.5 \log_{10} \frac{S}{S_0}, \tag{1.49}$$

where S counts the photons measured by a detector. That is,

$$S = \int \frac{f_\nu \mathrm{T}_\nu \mathrm{d}\nu}{\mathrm{h}\nu}, \tag{1.50}$$

where f_ν is the spectral flux density, T_ν the filter transmission function and h is Planck's constant. S_0 normalises the magnitudes so that the zero point corresponds to a bandpass-averaged spectral flux density of f_ν=3631 Jy. The magnitude of a galaxy as measured over different regions of the spectrum can be used to classify its colour. The mass of a galaxy is a further important property which can be estimated from observational data in a number of ways. Common approaches use either the known proportionality between luminosity and mass or the velocity dispersion measured within the galaxy in conjunction with basic gravitational theory. Additionally, absorption and emission lines in the spectra can indicate numerous things, including the types of stars contained within the galaxy and its star-formation history.

Galaxy redshift surveys have seen a dramatic improvement in recent years, yet there are still even more impressive surveys on the horizon. For example, the Large Synoptic Survey Telescope (LSST[4]), a wide-field survey with a 9.6 deg^2 field of view is scheduled for first light in 2021. Eventually LSST aims to observe

[4] `http://www.lsst.org/lsst/`

40 billion objects within a 30,000deg^2 field. The Euclid[5] space mission has also been recently confirmed and is scheduled for launch in 2020. The aim of Euclid is to accurately measure the acceleration of the Universe in order to better our understanding of dark energy and dark matter. It will measure the shapes and redshifts of galaxies and galaxy clusters out to $z \sim 2$. Euclid is set to provide deep images and spectra over at least half of the entire sky, providing a map of the Universe 75 times larger than what is currently available. Even though these surveys have not yet begun a significant amount of work is already ongoing to ensure that the framework and algorithms necessary to interpret the new data are already in place when the surveys commence.

1.4.2 Redshift-space distortions

One of the limitations of using redshift to infer distance is that a galaxy's velocity is not purely due to the expansion of the Universe. Distance measurements are subject to dynamical distortions as any additional velocity will introduce a Doppler shift to the measured spectrum which is non-trivial to disentangle from the redshift due to the expansion. That is, the Hubble equation becomes

$$v = H_0 d + v_{\rm pec}, \tag{1.51}$$

where $v_{\rm pec}$ is the (peculiar) velocity not associated with the expansion. We are also vulnerable to geometrical distortions if an incorrect cosmology is used to convert the redshifts into distances. Hence, distances inferred from redshifts are not 'true distances' and galaxies are said to be observed in 'redshift-space' rather than real-space, where true positions are known.

Dynamical redshift-space distortions are primarily due to two mechanisms. Firstly, the infall of objects towards high density regions due to gravity causes the galaxy distribution to appear thinner along the line of sight. This is often referred to as the Kaiser effect (Kaiser 1987). The motions are coherent, subtle and affect large-scales. Fig. 1.4 provides a cartoon explanation of the effect; overdense regions provide a gravitational force which acts against the Hubble expansion, leading to the region appearing contracted along the radial direction when viewed in redshift space. Conversely, underdense regions expand faster than

[5] `http://sci.esa.int/euclid/`

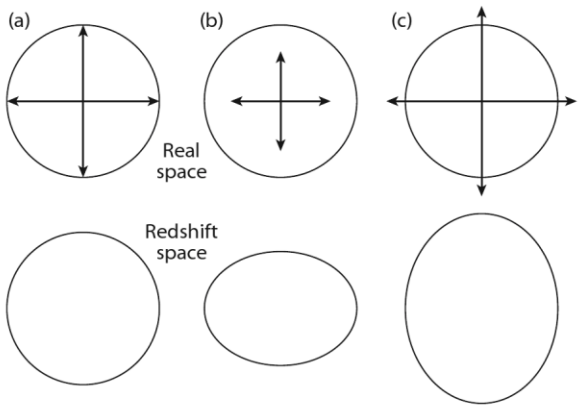


Figure 1.4 Illustration of redshift-space distortions due to the Kaiser effect. Top row represents real space and bottom row represents the shape of the region inferred in redshift space. Left column (**a**) represents a region at the mean density that follows the Hubble flow, leading to a correct inference of the radial width. Central column (**b**) represents an overdense region where the local expansion rate is smaller than the Hubble flow, leading to a smaller inferred width. Right column (**c**) represents an underdense region expanding faster than average. This leads to an overestimate of the radial width and the regions appear even more underdense in redshift space. (Image from Loeb & Furlanetto 2013).

average and appear elongated in the radial direction.

On smaller scales, galaxies also exhibit random peculiar motions. These motions tend to 'stretch out' clusters in redshift space producing long thin clusters of galaxies that appear to point directly back at the observer. These elongated structures are often called 'fingers of God' and have been seen in many galaxy surveys (e.g. Hawkins et al. 2003).

Whilst these distortions introduce an undesirable uncertainty into the measurement of galaxy distributions, they also provide an opportunity to test our cosmological models. For example, their statistical quantification has been shown to provide a robust method for measuring the growth of structure on very large scales (Percival et al. 2004, Cabré & Gaztañaga 2009, Samushia et al. 2012). Measuring this growth rate as a function of redshift can also constrain modified gravity models (Guzzo et al. 2008, Okada et al. 2013).

For the case of geometric distortions, if the wrong cosmological model is assumed when converting redshifts to distances then the inferred clustering will contain detectable distortions. This was first recognised by Alcock & Paczynski (1979) who noted that an incorrect cosmological model will have a different effect

for distances inferred along the line of sight than for angular distances. This led to the development of the Alcock-Paczyński test, which looks for statistical anisotropy as an indicator of an incorrect cosmological model; in a properly chosen cosmology, spheres will maintain a uniform ratio of line-of-sight to angular extent. These potential geometric distortions have been used by many authors to constrain cosmological parameters (Blake et al. 2012, Reid et al. 2012, Sutter et al. 2014).

1.4.3 N-body simulations

The increase in computational resources and technological progress of recent years has led to numerical simulations becoming an integral part of the field of cosmology. Theoretical modelling of non-linear scales is extremely difficult but numerical simulations allow us to simulate a universe in which the physics is known and follow its evolution. Although more complex simulations try to model baryonic physics, many simulations consider only dark matter interacting gravitationally. The field of numerical cosmology has been developing since the first astronomical N-body simulations were developed in the 1960s (von Hoerner 1960, Aarseth 1963). Notable modern dark matter simulations include Millennium[6], DEUS[7] and Bolshoi[8].

Such simulations model N 'particles' influenced by their mutual gravitational forces. The total force on a given particle will be the superposition of the forces due to all of the other particles. This net force allows the position and velocity of each particle at a time Δt later to be calculated. Depending on the scale and aims of the simulation, the individual particles can represent a large range of masses, from individual atoms to stars or larger. A significant limitation of this approach is that computational time scales as N^2, hence much work has been put into algorithmic and hardware developments to allow the size of numerical simulations to continuously increase, with N reaching the billions today.

In order to run these numerical simulations, one must chose the initial conditions – the starting redshift and initial positions and velocities of the particles. The standard approach is to assume Gaussian initial perturbations and

[6] `http://www.mpa-garching.mpg.de/galform/virgo/millennium/`
[7] `http://www.deus-consortium.org/`
[8] `http://hipacc.ucsc.edu/Bolshoi/`

use the Zel'dovich approximation introduced in Section 1.2.5 (Klypin & Shandarin 1983, Efstathiou et al. 1985) although alternative methods have been suggested (Scoccimarro 1998). The choice of initial conditions can also be refined using current observational data and the correct determination of our Universe's initial conditions may itself be the focus of the simulation (Wagner et al. 2010).

As well as simulating the distribution of individual particles and hence the underlying density field, it is useful to identify collapsed (gravitationally bound) structures which represent dark matter halos. The statistics of these structures can be compared to the theoretical models outlined in Section 1.2.1 and to observational data of the real Universe. Common methods of halo identification in numerical simulations include the Friends-of-Friends approach discussed above (Davis et al. 1985) and various techniques based on identifying local density maxima (Warren et al. 1992, Lacey & Cole 1994, Klypin & Holtzman 1997).

1.5 This thesis

The study of the cosmic web, as discussed in Section 1.3, is a relatively new and still developing field. Various classifications of the cosmic web within numerical simulations are discussed in the literature but fewer studies have attempted the more difficult task of classifying the web within the real Universe, as observed by galaxy redshift surveys. The recent development of new techniques, such as the tidal tensor prescription of Hahn et al. (2007), along with the availability of large, high-completeness observational galaxy catalogues has made such a classification of the observed cosmic web possible, providing an opportunity to analyse the largest structures in the known Universe.

The cosmic web is understood to exist as a product of huge tidal forces that have been deforming the cosmic density field since early epochs. As a result, the Universe exhibits significantly different regions covering a vast range of densities, shapes and sizes. It is a natural question, then, to ask how these different regions influence the structures that evolve within them.

The dependencies of galaxies and halos on their environment have been extensively studied for many years. It is well know that galaxy properties such as luminosity and colour are correlated with the mass of their host halo, and this has contributed significantly to theories of galaxy formation (Peacock & Smith 2000,

Moster et al. 2010). However, a full understanding of the role of the larger-scale environment during galaxy formation has not yet been achieved.

There is evidence that both galaxy properties and halo properties are correlated with their environments; recent studies have shown a dependence of galaxy colours and luminosities, among other properties, on environment (Blanton et al. 2005). Similarly, many halo properties including mass, spin, shape and assembly histories have been shown to exhibit environmental dependencies (Gao et al. 2005, Avila-Reese et al. 2005). However, there are a variety of ways to define environment. Often the definition of environment is derived from the local density field but this can still be computed on a range of scales. A complete picture of the extent of environmental influence requires the large-scale geometrical environments of the cosmic web to be considered in addition to the more standard tests based on the density field. Understanding these environmental dependencies is of high importance for many areas of astrophysics and cosmology, in particular, within galaxy formation models where the effect of the environment is often implicitly limited and when predicting the galaxy occupation statistics of halos where a dependence on halo mass alone is often assumed.

There is an increasing body of evidence suggesting that the cosmic web does influence the structures that evolve within it. Recent studies of the web within simulations claim to have found a correlation between halo properties and their web-environment (Zhang et al. 2009, Dubois et al. 2014, Metuki et al. 2015), with the strongest correlations seen for tensor properties such as spin. Similarly, recent works have investigated the correlation between galaxy properties and their web-environment within observational datasets, but these results are less clear and often do not disentangle the direct effects of the cosmic web from the indirect effects of the density field (Zhang et al. 2013, Darvish et al. 2014, Alpaslan et al. 2015 (in prep.)). Hence, further investigation is needed in order to develop a clear understanding of the extent of the influence of the cosmic web on LSS. This forms the motivation for this thesis, which aims to quantitatively detect any modulating effects of the quasilinear structures of the cosmic web on LSS.

Accordingly, the remainder of this thesis presents a number of investigations into the cosmic web: its detection within cosmological datasets and its influence on large-scale structure. The structure of this thesis is as follows:

Chapter 2 details an extraction of the cosmic web from the MDR1 dark matter simulation utilising the tidal tensor. The dependence of the mass function of dark-matter halos on their location within the simulated web is deduced empirically from the simulation and compared to linear theory predictions.

Chapter 3 presents the Galaxy And Mass Assembly (GAMA) spectroscopic redshift survey and details how the tidal tensor prescription can be extended to classify the cosmic web within observational datasets. The geometric environments of the GAMA galaxy and group catalogues are presented along with a thorough investigation into the robustness of the classifications.

In Chapter 4 these classifications are used to investigate the modulation of galaxy and group properties by the cosmic web within the GAMA survey. The direct dependence of the galaxy luminosity function, and of the group mass function, on geometric environment is measured.

Chapter 5 presents initial results of an investigation into the stellar assembly histories of GAMA galaxies within the cosmic web. The VESPA algorithm, used to extract star-formation histories from the GAMA spectra, is introduced and discussed. Various statistics to characterise formation time are presented and calculated within the different components of the web.

Finally, Chapter 6 summarises and discusses the results of this thesis with a consideration of the limitations and further work.

Chapter 2

Halo abundances within the simulated web

Environmental modifications of the properties of galaxies and galaxy systems is one of the central issues in the study of the formation of galaxies and large-scale structure (e.g. Abell 1965, Mo & White 1996, Peacock & Smith 2000, Berlind & Weinberg 2002, Berlind et al. 2003). The different observed degrees of clustering of galaxies of different types can be traced to the root cause that the formation of dark-matter halos is perturbed by the local matter density (Kaiser 1984, Cole & Kaiser 1989), and the variation of galaxy bias with scale and with redshift can be understood quantitatively in this way. This standard piece of cosmological theory has two distinct elements, which are the galaxy contents of different dark-matter halos, and how the properties of the halo population itself can vary with location within the overall density field. In the majority of published work, 'galaxy environment' refers purely to the former aspect, and is a shorthand for properties such as the mass of the halo hosting a given set of galaxies (e.g. Muldrew et al. 2012). This thesis is concerned with the larger-scale question – in effect defining environment via the density on scales of a few Mpc, rather than < 100 kpc. Independent of the extent to which different halos may generate different galaxy contents, a detailed understanding of the large-scale modulation of halo properties is a necessary preliminary. It has the virtue that it can be studied theoretically using N-body simulations of dark matter only. Furthermore, observational analogues of halos can be obtained directly via catalogues of galaxy groups, so in principle there is much that can be learned

about this aspect of cosmological structure without having to address the full complexity of galaxy formation.

This chapter presents an application of the tidal tensor prescription to simulated data in order to extract the components of the cosmic web and define geometric environments. The dependence of the mass function of dark-matter halos on their environment within the cosmic web of large-scale structure is then investigated. A dependence of the halo mass function on large-scale mean density is a standard element of cosmological theory, allowing mass-dependent biasing to be understood via the peak-background split. On the assumption of a Gaussian density field, this analysis can be extended to ask how the theoretical mass function depends on the geometrical environment: clusters, filaments, sheets and voids, as classified via the tidal tensor. In the remainder of this chapter, the predictions of this model are presented and compared to results deduced empirically from N-body simulations.

Section 2.1 introduces the simulated dataset used and describes the application of the tidal tensor prescription and the resulting simulated cosmic web. Section 2.2 discusses how Gaussian statistics can be used to make theoretical predictions regarding the cosmic web; Section 2.3 summarises some standard results concerning the conditional mass function of halos, and how this is modulated in regions of different large-scale overdensity; Section 2.4 contrasts these predictions with results from simulations and Section 2.5 presents and discusses the main conclusions of this chapter.

2.1 The simulated cosmic web

2.1.1 The MDR1 simulation

This thesis utilises numerical data from the MultiDark Run 1 (MDR1) dark matter N-body simulation (Prada et al. 2012). MDR1 simulates a $1\,(h^{-1}\mathrm{Gpc})^3$ cubic volume with a mass resolution of $m_p = 8.721 \times 10^9\,h^{-1}M_\odot$ in a ΛCDM cosmology with $(\Omega_m = \Omega_{dm} + \Omega_b, \Omega_\Lambda, \Omega_b, h, n, \sigma_8) = (0.27, 0.73, 0.0469, 0.7, 0.95, 0.82)$. The corresponding halo catalogue was compiled using a Friends-of-Friends algorithm on the $z = 0.1$ snapshot, yielding a minimum mass cut of $M_{\mathrm{min}} = 10^{11.5}\,h^{-1}M_\odot$.

2.1.2 The tidal tensor prescription

Although the cosmic web is clearly visible in all sufficiently detailed observed and simulated distributions of matter, its complexity and variety of scales, shapes, densities and dimensionality makes it nontrivial to quantify. As discussed in more detail in Section 1.3.1, a number of different approaches have been proposed and developed: minimal spanning tree methods have been used to detect filaments (Barrow et al. 1985, Alpaslan et al. 2014); topological methods based on Morse theory (Sousbie 2011) and morphological methods based on feature extraction techniques (Aragón-Calvo et al. 2010) and the watershed transform (Platen et al. 2007) have all been used to identify the full range of web components. Additionally, both the tidal tensor and the velocity shear tensor, with theoretical motivations drawn from Zel'dovich theory (Zel'dovich 1970), are able to produce good visual matches to the cosmic web (Forero-Romero et al. 2009, Hoffman et al. 2012). Recently methods utilising both velocity and density information, such as the ORIGAMI method of structure identification (Falck et al. 2012) which considers the folding of a 3D manifold in 6D phase space, have been successfully applied to simulations. However, many of these methods cannot be applied to observational data as they require information on the peculiar velocity of galaxies. Here, the tidal tensor prescription of Hahn et al. (2007) is chosen for its theoretical underpinnings and applicability to both simulated and observational datasets.

The tidal tensor prescription is in essence a stability criterion based on linear dynamics at each point in space. Each location is classified as a void, a sheet, a filament or a knot depending on whether structure is said to be collapsing in 0, 1, 2 or 3 dimensions respectively. This can be derived from knowledge of the gravitational potential field, Φ, using the tidal (or deformation) tensor, T_{ij}, defined as the matrix of second derivatives (the Hessian) of $\tilde{\Phi}$[1]:

$$T_{ij} = \frac{\partial^2 \tilde{\Phi}}{\partial r_i r_j}, \tag{2.1}$$

where the usual definition of the Newtonian potential has been normalised to

$$\tilde{\Phi} \equiv \Phi/(4\pi G\bar{\rho}) \tag{2.2}$$

[1]All spatial derivatives ∂_i are defined with respect to physical coordinates $\mathbf{r}$ and not comoving ones $\mathbf{q} \equiv \mathbf{r}/a$. This removes the proportionality factor a^2 from the relation between $\tilde{\Phi}$ and δ and makes the notation less cumbersome.

to make this relation simpler. The density contrast is then directly given by Poisson's equation as the trace of the Hessian of $\tilde{\Phi}$: $\delta = T_{11} + T_{22} + T_{33}$. The three real eigenvalues of the symmetric T_{ij} allow for the classification of space; the number of positive eigenvalues is equivalent to the dimension of the stable manifold at the point in question. In linear theory $-\nabla\tilde{\Phi}$ is also the displacement field, $\underline{f}$. Hence, $\underline{\nabla} \cdot \underline{f} < 0$ implies collapse, and a positive eigenvalue of $\hat{T}$ implies collapse along the corresponding eigenvector.

Whilst a positive but infinitesimally small eigenvalue implies that structure is collapsing, it may not reach non-linear collapse for a significant period of time, if ever. This leads to an overestimation of the number of collapsed dimensions and resulting classifications that are a poor match to the visual impression of the web. Hence, in order to account for the finite time of collapse, an eigenvalue threshold, $\lambda_{\rm th}$, is introduced as a free parameter of the tidal tensor prescription method of classifying geometric environments.

A given point in space is classified as belonging to one of the different elements of the cosmic web according to the number of eigenvalues of $\hat{T}$ above the threshold at that point. Thus, let $\lambda_1 \geq \lambda_2 \geq \lambda_3$ be the three eigenvalues of $\hat{T}$ and define:

- **Voids**: all eigenvalues below the threshold ($\lambda_1 < \lambda_{\rm th}$).
- **Sheets**: only one eigenvalue above the threshold ($\lambda_1 > \lambda_{\rm th}$, $\lambda_2 < \lambda_{\rm th}$).
- **Filaments**: two eigenvalues above the threshold ($\lambda_2 > \lambda_{\rm th}$,$\lambda_3 < \lambda_{\rm th}$).
- **Knots**: all eigenvalues above the threshold ($\lambda_3 > \lambda_{\rm th}$).

In this approach, the eigenvalue threshold is an arbitrary quantity. One might be tempted to set it to zero and classify based on the sign of the eigenvalues, but, as discussed, this is not very attractive from the point of view of gravitational collapse. We have seen that the density contrast is the sum of the eigenvalues, and we know that nonlinear collapse involves a linear contrast of order unity. It is therefore natural to consider $\lambda_{\rm th}$ of order unity; this choice is refined below.

A different prescription, the so-called V-web approach (Hoffman et al. 2012), uses instead the eigenvalues of the velocity shear tensor $\Sigma_{ij} \equiv -(\partial_i v_j + \partial_j v_i)/2\,H$. Although this prescription allows for a finer discrimination on very small scales, Σ_{ij} converges to $f_g\,T_{ij}$ in linear theory (where $f_g \equiv d\ln\delta/d\ln a$), and therefore similar results should hold. Note that, as seen in the following chapters of this

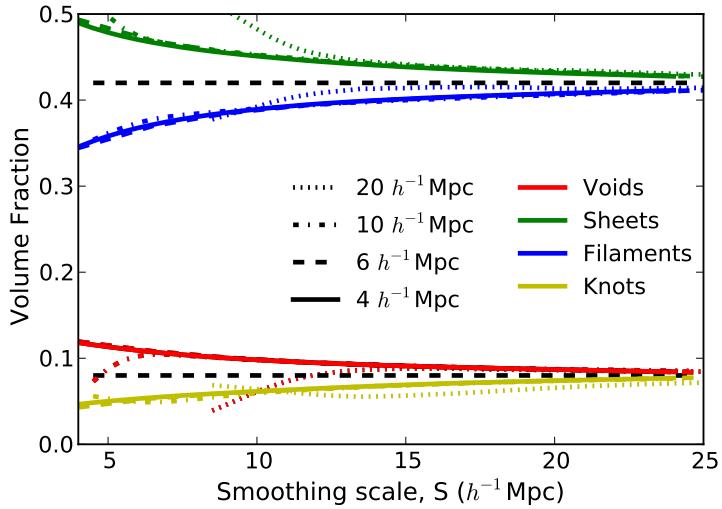


Figure 2.1 The fraction of the simulated volume which is classified as each of the four geometric environments, as a function of the width of the Gaussian used to smooth the density field, when a threshold of $\lambda_{\rm th} = 0$ is used. The different line styles show results for a variety of grid spacings, a, as shown by the key. Black dashed lines show the expected Gaussian asymptotes, as described in the text. All results tend towards the theoretical asymptotes at large smoothing scales. Results computed from different grid spacings converge at $a \approx 4h^{-1}$Mpc.

thesis, a natural extension is to apply this method to observational datasets, where the web would need to be defined using a redshift survey. If one works from the galaxy density, ρ_g, and uses a quasi-potential obeying $\nabla^2\tilde{\Phi} = -1 + \rho_g/\langle\rho_g\rangle$, the Hessian approach can still be used – but the velocity shear tensor is not observable.

The density field used to classify the different environments was computed by interpolating the dark-matter particle content on to a grid of spacing $a = 3.9\,h^{-1}$Mpc by cloud in cell interpolation, which uses multilinear interpolation to the eight nearest grid points to each galaxy. By investigating a variety of grid spacings, as shown in Fig. 2.1, it was found that results converge by $a \approx 4\,h^{-1}$Mpc.

In order for the tidal tensor to be well defined, the discrete density field must be smoothed. The purpose of this step is to suppress shot noise, and also to remove extreme non-linearities. The density field is smoothed with a Gaussian filter of variable width:

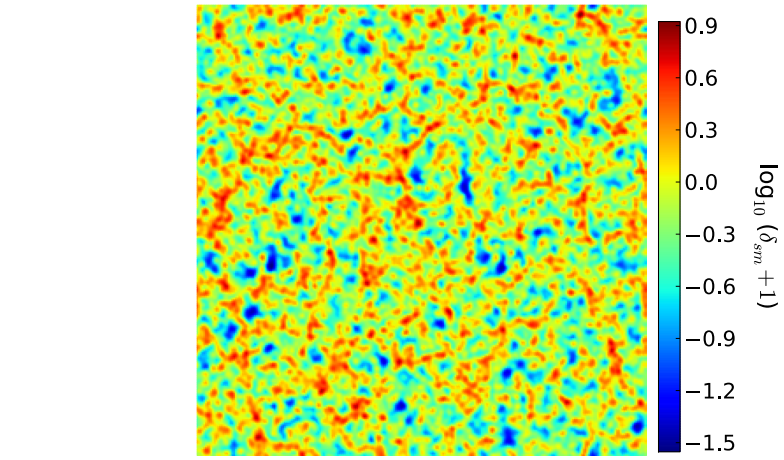
$$W(k) = \exp(-S'^2k^2/2); \tag{2.3}$$

the width of the Gaussian is denoted by S', since the common symbol σ is used elsewhere. Gaussian filtering is chosen in preference to spatial top-hat filtering for numerical stability. The results are closely equivalent to using a top-hat of radius $R \simeq \sqrt{5}S'$. The cloud in cell interpolation also inevitably introduces additional smoothing. By Taylor expanding the Fourier space window function for cloud-in-cell interpolation, one can show that this additional smoothing is approximately equivalent to smoothing with a Gaussian of width $S_{\rm c} = a/\sqrt{6}$. Hence, the effective smoothing scale is $S = \sqrt{S_c^2 + S'^2}$, and can be thought of as the typical length scale on which dynamical stability is being determined. This chapter presents results for two different filtering scales, $S = 4$ and $10\,h^{-1}$Mpc, chosen to investigate a range of scales in the quasi-linear regime. The deformation tensor was calculated at each point in the grid by solving Poisson's equation in Fourier space and then transforming back to configuration space (making extensive use of Fast Fourier Transforms throughout the process). Fig. 2.2a and Fig. 2.3a show a slice of the density field obtained for MDR1 after smoothing with the two chosen scales.

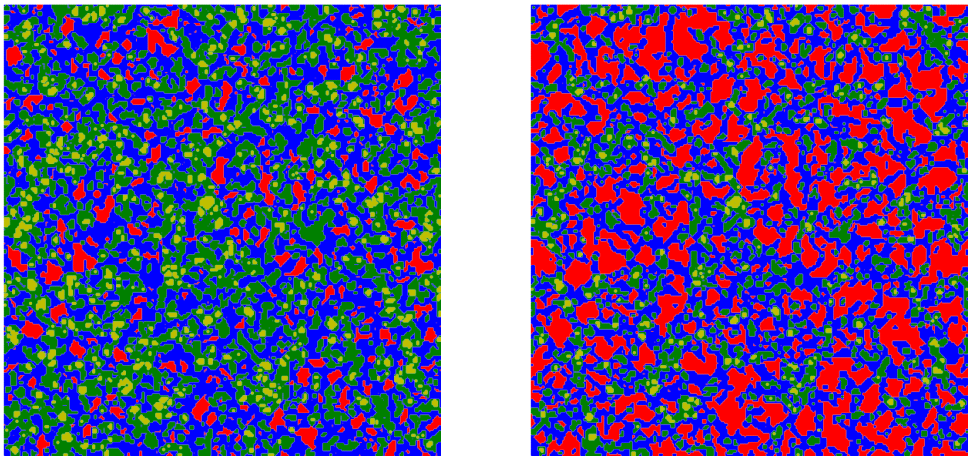
A useful consistency check is to look at the asymptotic volume fractions as the smoothing scale tends to very large values. For a threshold of $\lambda_{\rm th} = 0$, these expected asymptotes were derived by Doroshkevich (Doroshkevich 1970) and found to be 0.08 for knots and voids and 0.42 for sheets and filaments. Fig. 2.1 shows that the volume fractions are found to tend towards these expected values.

2.1.3 The eigenvalue threshold

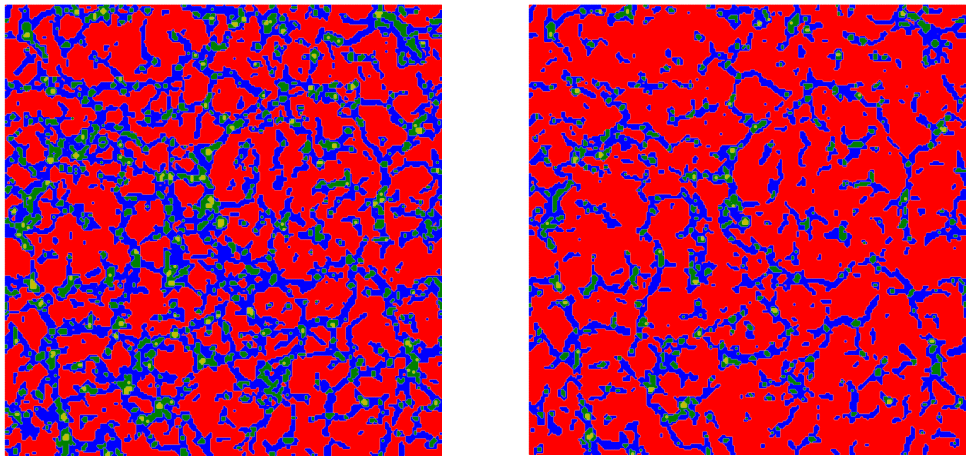
The value of the eigenvalue threshold to be used in this type of analysis is not precisely specified. One may choose $\lambda_{\rm th} = 0$ on the basis that this value discriminates between purely compressing or stretching tidal forces. Alternatively, $\lambda_{\rm th} = 1$ would correspond to complete 1D collapse, according to the Zel'dovich approximation, so a threshold of order unity is more natural. In any case, a threshold at $\lambda_{\rm th} = 0$ yields a low abundance of voids, compared to what one would expect from a visual examination of a large-scale structure map, and previous studies (e.g. Forero-Romero et al. 2009, Metuki et al. 2015) have chosen a threshold in order to yield a better match to the visually expected volume fractions. Fig. 2.2 and Fig. 2.3 show the resulting environments along a slice



(a) Smoothed density contrast field, $\delta_{\rm sm}$

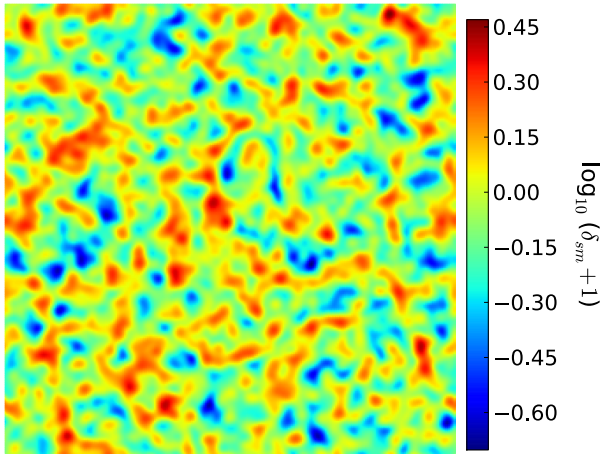


(b) $\lambda_{\rm th} = 0.0$ **(c)** $\lambda_{\rm th} = 0.2$

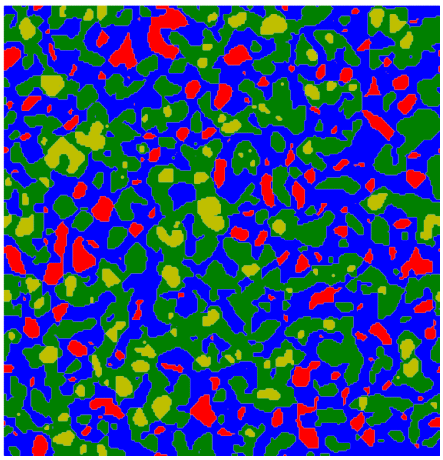


(d) $\lambda_{\rm th} = 0.4$ **(e)** $\lambda_{\rm th} = 0.6$

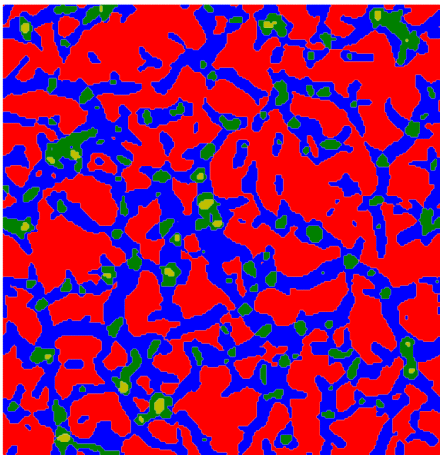
Figure 2.2 The environment classifications for a range of thresholds, $\lambda_{\rm th}$. **a:** The density contrast field smoothed with an effective smoothing scale of $4\ h^{-1}$Mpc. **b-e:** The resulting environment classifications computed from the smoothed density contrast field of **a** where the threshold, $\lambda_{\rm th}$, takes the values shown beneath each panel. **Colour Code:** Red: voids, blue: sheets, green: filaments, yellow: knots.



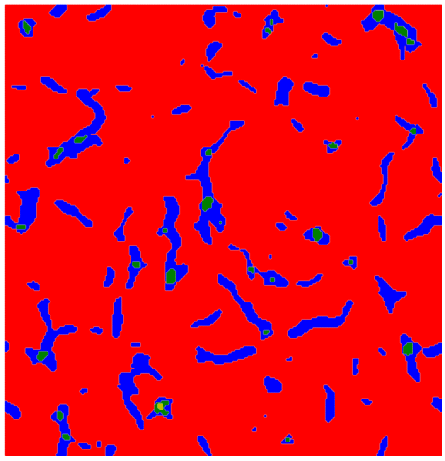
(a) Smoothed density contrast field, $\delta_{\rm sm}$



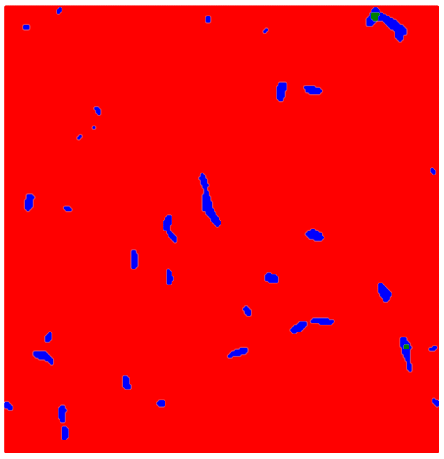
(b) $\lambda_{\rm th} = 0.0$



(c) $\lambda_{\rm th} = 0.2$



(d) $\lambda_{\rm th} = 0.4$



(e) $\lambda_{\rm th} = 0.6$

Figure 2.3 The environment classifications for a range of thresholds, $\lambda_{\rm th}$. **a:** The density contrast field smoothed with an effective smoothing scale of $10\ h^{-1}$Mpc. **b-e** The resulting environment classifications computed from the smoothed density contrast field of **a** where the threshold, $\lambda_{\rm th}$, takes the values shown beneath each panel. **Colour Code:** Red: voids, blue: sheets, green: filaments, yellow: knots.

of the MDR1 simulation, for a variety of thresholds, for each of the two chosen smoothing scales. It can be seen that the choice of threshold has a significant effect on the resulting definition of the cosmic web; as the threshold is increased from zero towards unity the voids become the dominant environment. As the threshold is increased further, eventually the full volume will be classified as a void, however, this is of course impractical and unphysical. One may argue that, for practical investigations, a threshold should be chosen such that it allows for the maximum amount of information to be extracted from the abundance of halos in these four environments. It is possible to use the value of the eigenvalue threshold to make sure that sufficient statistics are obtained for all of the different environments. In order to do so, the fraction of the total halo mass contained in each environment for different values of S and $\lambda_{\rm th}$, $FM_\alpha(S, \lambda_{\rm th})$, and the root mean square deviation of these fractions,

$$FM_{\rm rms}(S, \lambda_{\rm th}) = \left[\sum_\alpha (FM_\alpha - \overline{FM})^2/4\right]^{1/2}, \tag{2.4}$$

was computed. The optimal value of $\lambda_{\rm th}$ for each smoothing scale S is then the one that minimises $FM_{\rm rms}$. Besides this consideration, it is also important to make sure that the values of $\lambda_{\rm th}$ used yield physically sensible definitions for the different environments. One can then try to combine the aforementioned method with extra requirements, such as limiting the fraction of cells classified as voids that are overdense.

Fig. 2.4 shows the values of $FM_{\rm rms}$ for different smoothing scales and eigenvalue thresholds computed from the simulation. In view of this figure, two eigenvalue thresholds for each of the two smoothing scales were chosen:

$$S = 4\,h^{-1}{\rm Mpc} \longrightarrow \lambda_{\rm th} = 0.4,\, 0.6, \tag{2.5}$$

$$S = 10\,h^{-1}{\rm Mpc} \longrightarrow \lambda_{\rm th} = 0.1,\, 0.3. \tag{2.6}$$

In both cases the first value was chosen by restricting the fraction of overdense void cells to be smaller than 10%, while the second slightly larger threshold was chosen to illustrate how the results depend on this choice. The resulting environments computed with the first, optimal value of the threshold for the two smoothing scales are shown in Fig. 2.5. As expected, both slices display a

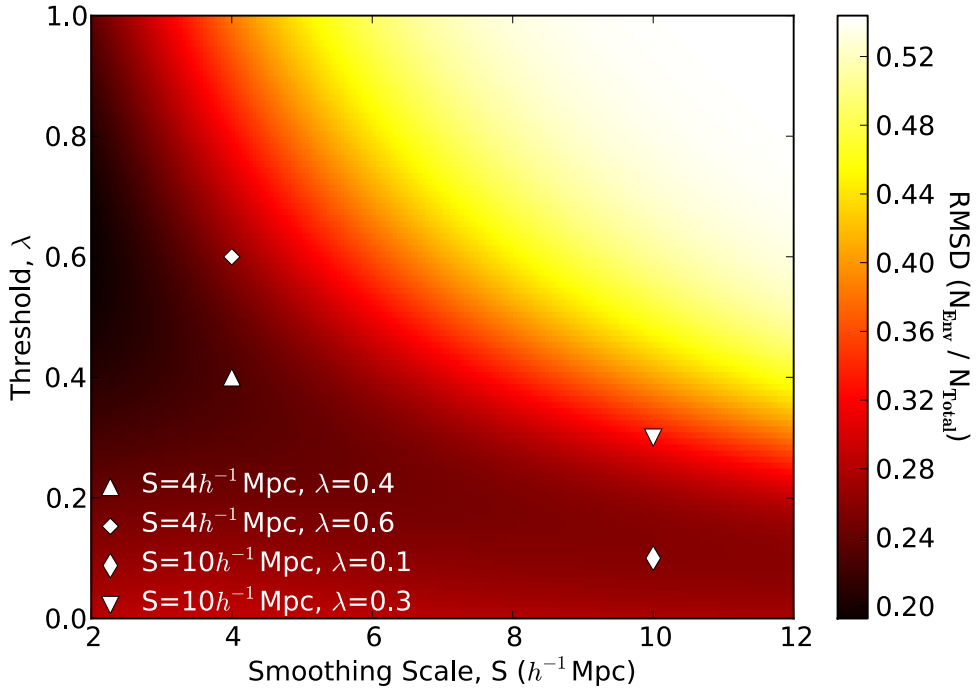
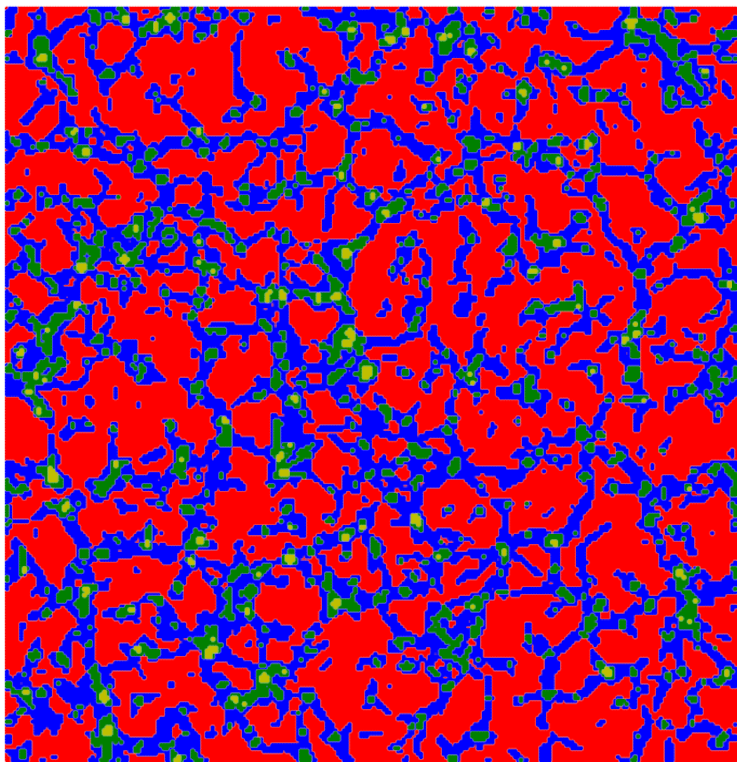
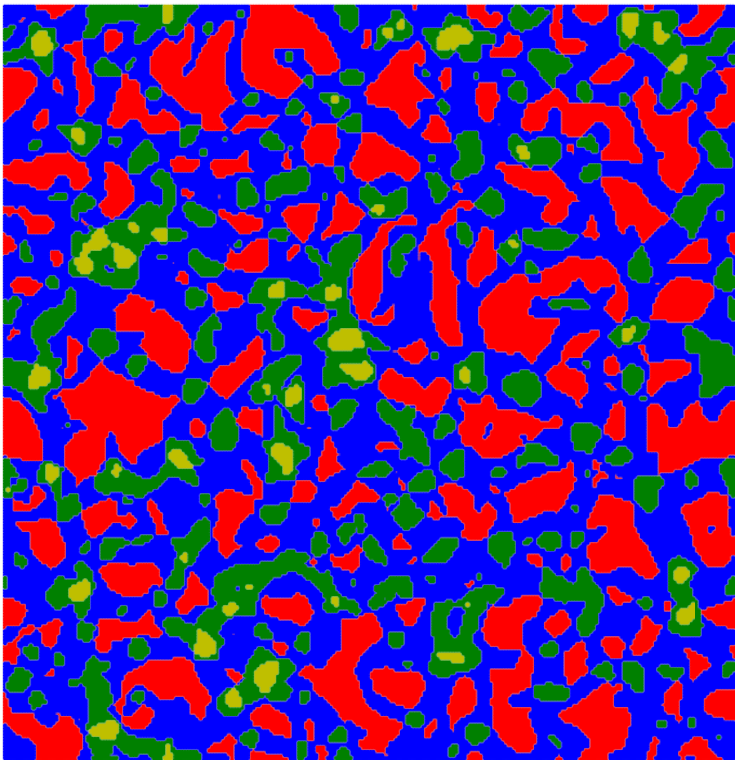


Figure 2.4 Illustrating how the partition of the cosmic mass density between different geometrical environments varies with smoothing scale and threshold. A practically useful partition will place approximately equal quantities of mass in the four environments, and the colour scale shows the dispersion in these mass fractions. The optimum is approximately $\lambda = 0.4$ for $S = 4\,h^{-1}\mathrm{Mpc}$ and $\lambda = 0.1$ for $S = 10\,h^{-1}\mathrm{Mpc}$ (shown as points); but an additional threshold for each smoothing scale is also considered, 0.6 and 0.3 respectively, to illustrate how the results depend on the choice of these parameters.



(a) $S = 4h^{-1}$Mpc, $\lambda_{\rm th} = 0.4$



(b) $S = 10h^{-1}$Mpc, $\lambda_{\rm th} = 0.1$

Figure 2.5 The environment classifications for the optimal choice of parameters: **a)** $S = 4h^{-1}$Mpc, $\lambda_{\rm th} = 0.4$ and **b)** $S = 10h^{-1}$Mpc, $\lambda_{\rm th} = 0.1$. **Colour Code:** Red: voids, blue: sheets, green: filaments, yellow: knots.

similar basic skeleton, with the large smoothing scale environments finding larger structures. We will now look at what the theory can predict, using this definition of the cosmic web, focusing especially on the population of dark matter halos in different web environments.

2.2 Gaussian statistics in the cosmic web

It is interesting to study the statistics of the cosmic web in the limit that $\tilde{\Phi}$ is a Gaussian random field. It is also assumed that the potential is smoothed on a scale R_e with some window function $W_e(kR_e)$. In this case the probability density function of the eigenvalues λ_i is given by:

$$dP(\lambda_i) \equiv p(\rho, \theta, \nu)\, d\rho\, d\theta\, d\nu \tag{2.7}$$

$$\equiv 225\sqrt{\frac{5}{2\pi}}\rho\,(\rho^2 - \theta^2)\,\exp\left[-\frac{1}{2}(15\rho^2 + 5\theta^2)\right] \times \frac{e^{-\nu^2/2}}{\sqrt{2\pi}}\, d\rho\, d\theta\, d\nu, \tag{2.8}$$

where $\nu \equiv (\lambda_1 + \lambda_2 + \lambda_3)/\sigma_{ee} \equiv \delta_e/\sigma_{ee}$, $\rho \equiv (\lambda_1 - \lambda_3)/2\sigma_{ee}$ and $\theta \equiv (\lambda_1 - 2\lambda_2 + \lambda_3)/2\sigma_{ee}$. The derivation of this result is outlined in Appendix A.1 and has been widely used in the literature (Doroshkevich 1970, Bardeen et al. 1986, Bond & Myers 1996, Rossi 2013). Note that these results are usually quoted in terms of the ellipticity e and prolateness p which are related to the stated variables by $e \equiv \rho/\nu$, $p \equiv \theta/\nu$.

The restriction $\lambda_1 \geq \lambda_2 \geq \lambda_3$ implies $\rho \in [0, \infty)$, $\theta \in [-\rho, \rho]$, and selecting a given type of environment further constrains the dynamical range of ν. Let $\lambda_{\rm th}$ be the eigenvalue threshold used for the classification, and let α denote the number of eigenvalues above this threshold for each case. Then the integration

limits for ν are $f_1^\alpha(\rho,\theta) < \nu - \nu_{\rm th} < f_2^\alpha(\rho,\theta)$, with $\nu_{\rm th} \equiv 3\lambda_{\rm th}/\sigma_{ee}$ and

$$f_1^\alpha(\rho,\theta) = \begin{cases} -\infty, & \alpha = 0 \text{ (voids)} \\ -3\rho - \theta, & \alpha = 1 \text{ (sheets)} \\ 2\theta, & \alpha = 2 \text{ (filaments)} \\ 3\rho - \theta, & \alpha = 3 \text{ (knots)} \end{cases}, \tag{2.9}$$

$$f_2^\alpha(\rho,\theta) = \begin{cases} -3\rho - \theta, & \alpha = 0 \text{ (voids)} \\ 2\theta, & \alpha = 1 \text{ (sheets)} \\ 3\rho - \theta, & \alpha = 2 \text{ (filaments)} \\ \infty, & \alpha = 3 \text{ (knots)} \end{cases}. \tag{2.10}$$

The volume fraction in environments of type (α, $\nu_{\rm th}$) can be calculated as the probability for a random point in space to belong to that type:

$$F_V(\alpha,\nu_{\rm th}) \equiv \int_0^\infty d\rho \int_{-\rho}^{\rho} d\theta \int_{\nu_{\rm th}+f_1^\alpha(\rho,\theta)}^{\nu_{\rm th}+f_2^\alpha(\rho,\theta)} d\nu\, p(\rho,\theta,\nu). \tag{2.11}$$

Notice that the volume fraction is a universal function of the normalised threshold $\nu_{\rm th}$. These volume fractions, as well as the fraction of mass in each environment, as a function of $\lambda_{\rm th}$ are shown in Fig. 2.6 and Fig. 2.7 respectively. As can be seen, the eigenvalue thresholds chosen in the previous section try to maximise statistics of the four environments simultaneously. The figure also shows the fractions measured from the MDR1 simulation. The match is rather good for the larger value of the smoothing scale. With this Gaussian framework in place, the next step is to look at the halo mass function in the different environments.

2.3 The conditional halo mass function

Most of the information about the non-linear accretion of dark matter halos is encoded in the mass function $n(M)\,dM$, the comoving number density of halos with mass $M \in (M, M + dM)$. As discussed in more detail in Section 1.2.2, Press & Schechter (1974) (PS hereafter) developed a theoretical framework for calculating the mass function based solely on Gaussian statistics and non-linear spherical collapse. Their main result is that the collapsed fraction $F(> M)$ (the fraction of the Universe collapsed into virialized structures of masses larger than

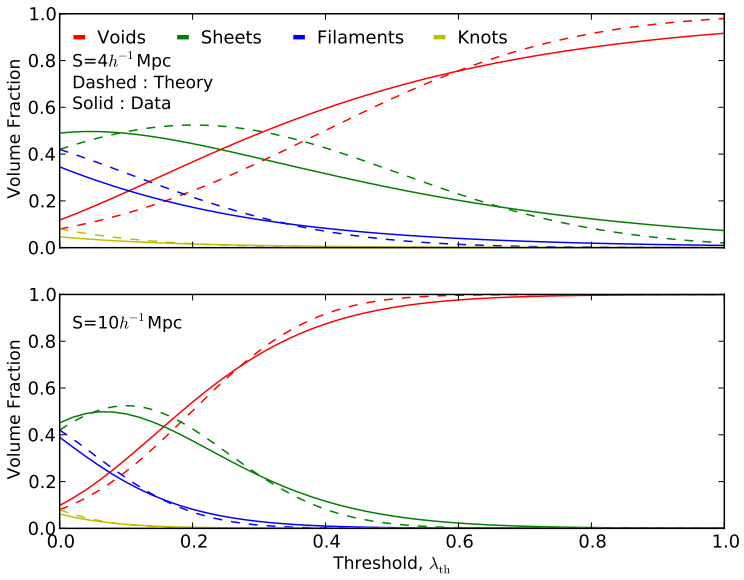


Figure 2.6 The Gaussian-field prediction for the volume fractions in the four environments for $S = 4,\, 10\,h^{-1}\mathrm{Mpc}$ as a function of the eigenvalue threshold (dashed lines). The volume fractions measured from the MDR1 simulation (solid lines) are also shown for comparison. The results are shown for voids (red), sheets (green), filaments (blue) and knots (yellow)

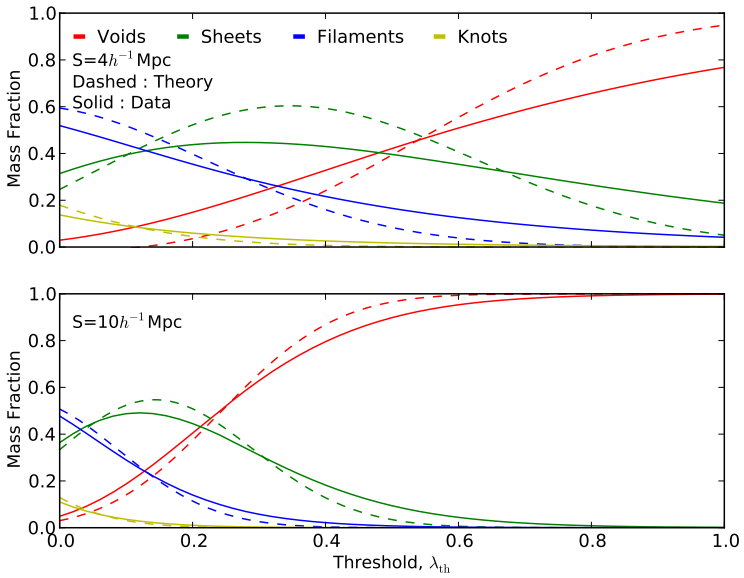


Figure 2.7 The Gaussian-field prediction for the fraction of mass in each of the four environments for $S = 4,\, 10\,h^{-1}\mathrm{Mpc}$ as a function of the eigenvalue threshold (dashed lines). The mass fractions measured from the MDR1 simulation (solid lines) are also shown for comparison. The results are shown for voids (red), sheets (green), filaments (blue) and knots (yellow)

M) can be written as a universal (cosmology independent) function of the variable

$$\nu_c \equiv \frac{\delta_c}{\sigma(M)}, \tag{2.12}$$

where $\delta_c \simeq 1.686$ is the linear threshold for spherical collapse and $\sigma^2(M)$ is the variance of the density contrast filtered on a scale $R = (3M/4\pi\rho_0)^{1/3}$, given in linear perturbation theory by

$$\sigma^2(M) = \frac{1}{2\pi^2} \int_0^\infty P_k \, |W(kR)|^2 \, k^2 \, dk, \tag{2.13}$$

where $W(kR)$ is the transform of a top-hat window function and P_k is the matter power spectrum. The PS result is

$$F(> M) = 1 - \mathrm{erf}\left(\frac{\nu_c}{\sqrt{2}}\right), \tag{2.14}$$

from which the mass function can be calculated as

$$n(M)\, dM = \frac{\rho_0}{M} \left| \frac{dF(> M)}{d \ln M} \right| d \ln M. \tag{2.15}$$

Note that it is often convenient to express this in the form of the multiplicity function: $M^2 n(M)/\rho_0$, which is the fraction of the mass of the Universe carried by halos in unit range of $\ln M$.

2.3.1 The excursion set formalism

The PS result was given a more solid mathematical foundation by Bond et al. (1991) and by Lacey & Cole (1993) through the so-called excursion set formalism. In this framework the density contrast at a given point as a function of the smoothing scale, $\delta(R)$, forms a random walk in $\delta - R$ space. The collapsed fraction can then be calculated as the fraction of all those random walks that, starting at $\delta(R \to \infty) \to 0$, make a first crossing above the δ_c threshold at some $R > R(M)$. In its original form, this formalism has one important caveat: for the walks to be completely random (i.e. with uncorrelated steps) the density contrast must be filtered using a top-hat window function in k-space, whereas the original PS result assumed top-hat filters in configuration space. There exist

a few approaches in the literature to take into account the non-zero correlations between steps when using different window functions (Peacock & Heavens 1990, Maggiore & Riotto 2010, Musso et al. 2012).

The power of the excursion set formalism is that it can be extended to study the merging history of halos as well as the conditional distribution of halos in environments with different large-scale densities (Mo et al. 1996), since both problems can be addressed by studying random walks crossing two barriers of different heights. Assuming Gaussian statistics one can compute the conditional probability for the density contrast smoothed over R_h to have a value δ_h given the value δ_e when smoothed on a scale R_e (here the subscripts e is used for environmental quantities and h for halos):

$$P(\delta_h|\delta_e)\,d\delta_h = \frac{d\nu_h}{\sqrt{2\pi(1-\varepsilon^2)}}\exp\left[-\frac{1}{2}\frac{(\nu_h-\varepsilon\,\nu_e)^2}{1-\varepsilon^2}\right], \tag{2.16}$$

where $\nu_e \equiv \delta_e/\sigma_{ee}$, $\nu_h \equiv \delta_h/\sigma_{hh}$, $\epsilon \equiv \sigma_{eh}^2/(\sigma_{ee}\,\sigma_{hh})$ and the covariance is defined as

$$\sigma_{ab}^2 \equiv \frac{1}{2\pi^2}\int_0^\infty P_k\,W_a(kR_a)\,W_b^*(kR_b)\,k^2dk. \tag{2.17}$$

Notice that different window functions may have been used for δ_h and δ_e. Equation (2.16) is in fact the same result found in the unconditional case with a rescaling of the variable

$$\nu_h \to \frac{\nu_h-\varepsilon\,\nu_e}{\sqrt{1-\varepsilon^2}}, \tag{2.18}$$

and therefore the same rescaling applies to Equation (2.14) to obtain the conditional collapsed fraction

$$F(>M|\delta_e) = 1-\mathrm{erf}\left(\frac{\nu_c-\varepsilon\,\nu_e}{\sqrt{2\,(1-\varepsilon^2)}}\right). \tag{2.19}$$

In both this and the unconditional problem, the random-walk approach solves the 'cloud-in-cloud' issue and yields a collapse fraction that is simply twice the area under the tail of the one-point density distribution; again, this simple result does not hold in the case of correlated trajectories. There exist other approaches in the literature that improve on this simple formalism to take into account the correlation between excursion steps as well as the peak nature of collapsed structures (Paranjape et al. 2013). These are not considered here, since none of

the trajectory-based approaches match 'exact' numerical results.

The original PS mass function in particular is well known to be a poor fit to N-body data, and more precise fitting formulae have been developed. These retain the main element of the PS philosophy by continuing to write the mass function with a universal dependence on the variable ν_c. In general, this works extremely well, although in detail slight deviations from this universal scaling do exist (e.g. Sheth & Tormen 1999, Jenkins et al. 2001, Peacock 2007, Tinker et al. 2008, Watson et al. 2013). A more complex problem is what to do about the conditional mass function. Since the PS form for this uses the unconditional mass function with a change of variable from ν_c to $\nu_{\rm eff} \equiv (\nu_c - \varepsilon\,\nu_e)/\sqrt{1-\varepsilon^2}$, it is tempting to use the 'observed' functional dependence of the mass function on ν_c, substituting it by $\nu_{\rm eff}$. As shown explicitly in Section 2.4.2, this rescaling seems to be a good approximation only for large smoothing scales and mild environmental overdensities, and it breaks down for large values of δ_e. This comes as no surprise, since the same approach for the conditional mass function as a means to study the merging histories of halos has been shown to fail to match N-body data (Cole et al. 2008).

As with the original PS approach, the excursion set method provides a framework for relating the linearised Gaussian density field to the abundance of virialized objects that form in the fully non-linear density. For this reason, the environmental density entering Equation (2.19) must be not the value of the local physical density Δ_e, but its linear extrapolation, which can be estimated using the prescription given by Bernardeau (1994):

$$1 + \Delta_e = \left(1 - \frac{\delta_e}{\delta_c}\right)^{-\delta_c}. \tag{2.20}$$

2.3.2 The effective-universe approach

The problem of the conditional mass function can also be approached from a different point of view. A well-known result of gravitational theory (which holds both in general relativity and in the Newtonian limit) is that any spherically symmetric system behaves, at a fixed radius, in a manner equivalent to a spherical sub-region of a homogeneous universe with some effective cosmological parameters. Thus, an overdense spherical region embedded in a Friedmann-Lemaître-Robertson-Walker (FLRW) universe will have a radius-time history

identical to that of a different effective FLRW model (Weinberg 1972, Peebles 1993).

According to this result, and neglecting any error involved in applying it to non-spherical systems, one should be able to calculate an accurate conditional mass function for any degree of environmental overdensity, by exploiting the fact that the fitting formula for the unconditional mass function is claimed to be universal. Given a set of effective cosmological parameters, one should then be able to calculate the mass function in the effective universe – which is then equivalent to calculating the desired conditional mass function for environments of the same density. This approach has been discussed in previous works (Gottlöber et al. 2003, Goldberg & Vogeley 2004), and in fact has been shown to be equivalent to the excursion set method in the limit of large smoothing scales (Martino & Sheth 2009). The main steps of this method are outlined here, and Appendix A.2 gives a more detailed description.

Since the small-scale perturbations inside the effective universe have their origin in the same primordial power spectrum as perturbations in the background FLRW model, the variance of the linear overdensity field in the effective universe is just a rescaled version of the variance outside:

$$\sigma_{\mathrm{eff}}(M) = D_g\, \sigma(M, R_e). \tag{2.21}$$

Here D_g is the ratio of the growth factor in the effective universe to the growth factor in the background cosmology, and $\sigma(M, R_e)$ is the variance of the density field in the background corrected for the size of the environment (see Appendix A.2, in particular Eq. A.14, for further details). Thus, according to the PS theory, the mass function in this effective universe has the same functional form as the universal mass function Eq. 2.14, using the rescaled variance in Eq. 2.21.

2.3.3 Dependence on the environmental density

The aim is to give a prediction for the abundance of halos of different mass in different types of environment. In the excursion set approach, this abundance is given by the number of first upcrossings of the collapse threshold by the density contrast field smoothed over a scale R_h, corresponding to the halo mass. Therefore we first need to study the joint probability for the density contrast δ_h smoothed

over this scale and the environmental tidal tensor eigenvalues, filtered on a scale R_e. As shown in Appendix A.1, this distribution is:

$$\frac{dP(\delta_h,\lambda_i)}{d\nu_h\,d\nu_e\,d\rho\,d\theta}=\frac{p(\rho,\theta,\nu_e)}{\sqrt{2\pi(1-\varepsilon^2)}}\exp\left[-\frac{(\nu_h-\varepsilon\,\nu_e)^2}{2\,(1-\varepsilon^2)}\right], \tag{2.22}$$

where $\nu_e\equiv\delta_e/\sigma_{ee}$.

As a first step, we are interested in the abundance of halos within environments of type $(\alpha,\nu_{\rm th})$ with local environmental density contrast $\delta_e/\sigma_{ee}\in(\nu_e,\nu_e+d\nu_e)$. In this case, integration over ρ and θ finds the joint distribution

$$\begin{aligned}P(\nu_h,\nu_e,\alpha,\nu_{\rm th})\,d\nu_e\,d\nu_h=&\frac{d\nu_e}{\sqrt{2\pi}}\,C(\nu_e)\,e^{-\nu_e^2/2}\\&\times\frac{d\nu_h}{\sqrt{2\pi(1-\varepsilon^2)}}\exp\left[-\frac{(\nu_h-\varepsilon\,\nu_e)^2}{2\,(1-\varepsilon^2)}\right].\end{aligned} \tag{2.23}$$

Here

$$C(\nu_e)\equiv 225\sqrt{\frac{5}{2\pi}}\iint_{\mathcal{A}(\nu_e)}d\rho\,d\theta\,\rho\,(\rho^2-\theta^2)\,e^{-\frac{1}{2}\left(15\rho^2+5\theta^2\right)} \tag{2.24}$$

and $\mathcal{A}(\nu_e)$ is the region in the plane (θ,ρ) defined by the conditions

$$\begin{aligned}&\rho\geq 0,\quad -\rho\leq\theta\leq\rho,\\&f_1^\alpha(\rho,\theta)\leq\nu_e-\nu_{\rm th},\leq f_2^\alpha(\rho,\theta).\end{aligned} \tag{2.25}$$

On the other hand, the probability of finding such an environment is given by

$$P(\nu_e,\alpha,\nu_{\rm th})\,d\nu_e=\frac{d\nu_e}{\sqrt{2\pi}}\,C(\nu_e)\,e^{-\nu_e^2/2}, \tag{2.26}$$

which cancels out when computing the conditional probability

$$P(\nu_h|\nu_e,\alpha,\nu_{\rm th})\,d\nu_h=P(\delta_h|\delta_e)\,d\delta_h, \tag{2.27}$$

where $P(\delta_h|\delta_e)$ is given in Eq. 2.16.

From this it is possible to extract a key general prediction of the Gaussian formalism: the only environmental quantity that determines the abundance of halos is the local density $\delta_e\equiv\lambda_1+\lambda_2+\lambda_3$. This conclusion arises from the fact that the treatment is based on the spherical top-hat collapse, which

disregards all couplings of the halo orientations with other combinations of the eigenvalues of the tidal tensor (i.e. the only non-zero, non-diagonal element of the covariance matrix is between ν_h and ν_e). While this assumption of zero coupling of gravitational collapse to tidal forces could be challenged in detail, it is striking that the geometrical environment is not predicted to have a more direct influence on the properties of halos at a given overdensity.

In the remainder of this chapter, this result is subjected to the test of confrontation with numerical simulations. Any evidence for an explicit dependence on geometrical environment would be interesting as it would relate to the issue of halo 'assembly bias' (e.g. Gao & White 2007). This term stands for the possibility that halo properties have some dependence on parameters beyond the local overdensity; the concept may apply either to the final-state properties of the halo or to its merger history (which potentially influences the galaxy contents of the halo). This issue is certainly under active consideration from the point of view of the observational dependence of galaxy properties on tidal forces (e.g. Yan et al. 2013). It is worth noting that this analysis is focused on overall halo abundances and not intrinsic halo properties, and hence it does not directly address the problem of assembly bias.

2.3.4 The four mass functions

Using the key result of the previous subsection, any halo statistic conditional to a given type of environment can be calculated within the excursion set formalism as the average of that statistic conditional to a given background density in that environment. E.g. for the conditional collapsed fraction:

$$F(>M|\alpha,\nu_{\rm th}) = \int_0^\infty d\rho \int_{-\rho}^{\rho} d\theta \int_{\nu_{\rm th}+f_1^\alpha(\rho,\theta)}^{\nu_{\rm th}+f_2^\alpha(\rho,\theta)} d\nu_e \times \frac{p(\rho,\theta,\nu)\, F(>M|\delta_e \equiv \sigma_{ee}\nu_e)}{F_V(\alpha,\nu_{\rm th})}, \tag{2.28}$$

where $F(>M|\delta_e)$ is obtained through either of the methods outlined in Section 2.3.

2.4 Comparison with simulations

2.4.1 Conditional density distributions

According to the theoretical framework described in Sections 2.2 and 2.3, the halo mass function should be the same in environments with the same density, depending only implicitly on the environment classification due to the different distribution of densities for each environment type. Therefore, the ability to predict the abundance of halos in each element of the cosmic web depends, on the one hand, on reproducing these distributions correctly and, on the other, on the accuracy of the model for the conditional mass function. These distributions are given, according to this formalism, by

$$P(\delta_e|\alpha,\nu_{\rm th})d\delta_e \equiv \frac{P(\nu_e,\alpha,\nu_{\rm th})}{F_V(\alpha,\nu_{\rm th})}d\nu_e, \tag{2.29}$$

where $P(\nu_e,\alpha,\nu_{\rm th})$ is given by Eq. 2.26.

This quantity is shown in Fig. 2.8 as measured from the MDR1 simulation for different values of S and $\lambda_{\rm th}$. The two top panels in this figure show how the density distributions measured directly from the simulation (continuous histogram) are very poorly described by the Gaussian theory, especially for high-density environments. This is not so surprising, since it is well known that the matter density is significantly nonlinear on the filtering scales used in this work. In order to understand these differences better, these results have been compared with the predictions of the lognormal distribution (Coles & Jones 1991), which has often been used as a convenient approximate model for the nonlinear density field. Specifically, the real density field was used to perform the environment classification, and then the distribution within each environment of the field that results from undoing the lognormal transformation was studied. This is given in terms of the real overdensity δ_r and its variance σ_r^2 as

$$\delta_G = \ln\left[(1+\delta_r)\sqrt{1+\sigma_r^2}\right]. \tag{2.30}$$

The distribution of this 'de-lognormalised' overdensity is shown in the bottom panels of Fig. 2.8. We can see that this field follows the Gaussian theoretical distribution much better for the largest smoothing scale $S = 10\,h^{-1}$Mpc. For smaller scales, however, the lognormal transformation is no longer a good

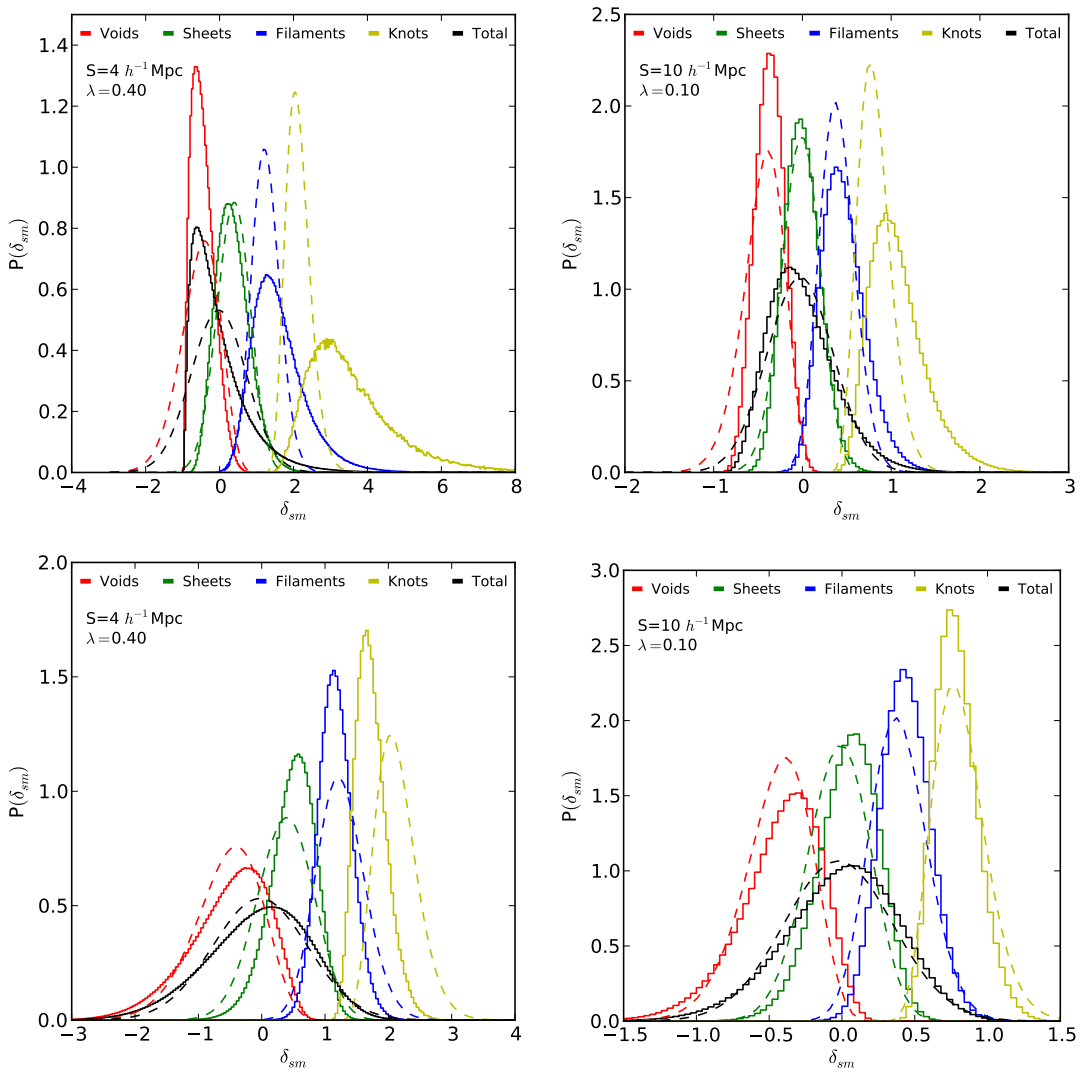


Figure 2.8 Overdensity distributions in each of the four environments and the overall overdensity distribution for $(S, \lambda_{\rm th}) = (4\,h^{-1}{\rm Mpc}, 0.4)$ (left panels) and $(S, \lambda_{\rm th}) = (10\,h^{-1}{\rm Mpc}, 0.1)$ (right panels). The dashed lines show the Gaussian theoretical prediction, while the solid histograms show the distributions extracted from the MDR1 simulation. In the top panels these histograms correspond to the distribution of the real density field, while the bottom panels show the distribution of the 'de-lognormalised' overdensity (see Eq. 2.30). The colour code is voids (red); sheets (green); filaments (blue); knots (yellow); overall distribution (black). In the same order, the density distribution for the 4 environments peaks on increasing values of δ_{sm}.

description of the non-Gaussianity of the density field, a result that has been reported by other authors (Kitaura et al. 2010).

2.4.2 Halo abundances

The overall multiplicity function in voids, sheets, filaments and knots is calculated by averaging $f(M|\delta_e)$ over δ_e using the overdensity distribution of each environment. As shown in the previous section, the Gaussian prediction for the density distribution $P(\delta_e|\alpha, \nu_{\rm th})$ is not a good description of the real distribution in most cases, even after attempting to account for non-Gaussianity using the lognormal approximation. Hence, even if the model for the conditional mass function were accurate, it would still not be able to predict $f(M|\alpha, \nu_{\rm th})$ correctly. For this reason, in order to isolate any inaccuracies due to the modelling of the conditional mass function from those due to the Gaussian approximation, the actual density distributions measured from the MDR1 simulation (solid lines in Fig. 2.8) were integrated over in order to obtain a theoretical prediction for the four mass functions. These are shown in Fig. 2.9 for the cases quoted in Eq. 2.6 together with the theoretical predictions for the excursion set and the effective universe approaches. These predictions are based on rescaling the universal collapsed fraction, which was estimated using the fitting formula proposed by Peacock (2007). The results are in qualitative agreement with Hahn et al. (2007), who used an eigenvalue threshold $\lambda_{\rm th} = 0$. As is discussed in the next section, the excursion set formalism is only able to make reasonable predictions for environments involving small overdensities (i.e. voids and sheets) and for large filter scales, while the effective-universe approach shows an overall better agreement with the data. Nevertheless, neither model is able to describe the data accurately. For the present, this work shows only an approximate understanding of the trends in halo properties with environment; accurate work will require calibration from numerical simulation, just as with the original PS mass function.

2.4.3 Universality of density dependence

The above results show that the excursion set approach is able to make relatively good predictions for large smoothing scales and mild environmental densities, but that it fails to do so for smaller values of S and high δ_e. This is a reasonable

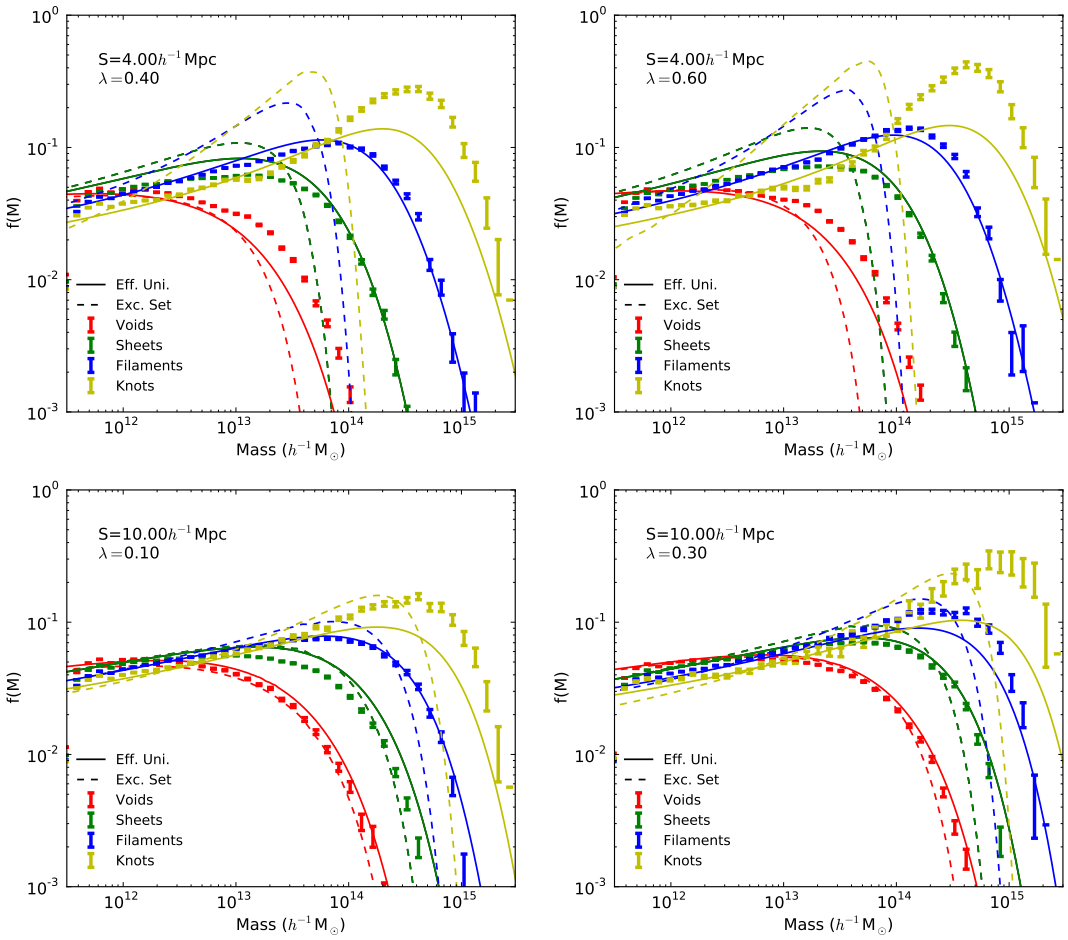


Figure 2.9 Multiplicity function for the four environments for the fiducial smoothing scales and eigenvalue thresholds. Solid and dashed lines show the theoretical predictions of the effective-universe approach and the standard excursion set model respectively, data points show the multiplicity functions measured from the MDR1 simulation. In increasing order of amplitude on high masses, the different multiplicity functions correspond to voids, sheets, filaments and knots. Reasonable agreement is only obtained for large smoothing radii and small $\lambda_{\rm th}$. This can be attributed to the fact that the standard prediction for the conditional mass function is only valid for densities below the collapse threshold and masses below the filter scale. The theoretical prediction was made by rescaling the empirical formula in Peacock (2007) for the mass function.

result: the excursion set model is based on following trajectories in the $\delta - R$ plane that, starting at some (R_e, δ_e), cross the threshold δ_c at some scale $R(M) < R_e$. In this regime the excursion set predicts that too many small-mass halos have already merged into larger ones, due to the fact that the large environmental density makes gravitational collapse more efficient. At the large-mass end, on the other hand, too few halos have formed, since the total halo mass is limited by the mass that can be found within the smoothing radius. Also, for small R_e the correlation between adjacent steps, which is generally ignored, may play a significant role, since the scale of the halo mass can be close to the scale of the environment. As can be seen from Fig. 2.9, in these limits, the effective-universe approach outperforms the excursion set, providing a better description for the conditional mass function – although far from per-cent level precise.

Nevertheless, one of the most important predictions of this formalism is that the abundance of halos should be the same for all environments with the same background density. Regardless of whether or not the theoretical prediction for the conditional mass function is quantitatively precise, the qualitative validity of this result can be tested with the simulated data.

Fig. 2.10 shows the multiplicity function,

$$f(M) \equiv |d\,F(< M)/d\,\ln M|\,, \tag{2.31}$$

for halos residing in the four different environments with restricted local overdensities, together with the environment-independent theoretical predictions derived from the excursion set (dashed lines) and the effective universe approach (solid lines) as detailed in Section 2.3. The values of the local overdensity were chosen to guarantee the simultaneous presence of as many different environments as possible. For the range of masses, smoothing scales and densities explored, it can be seen that the prediction that the abundance of halos should depend only on the environmental density and not on the environment classification holds very well, with little or no deviation within statistical errors. This agreement has been quantified as follows: for each overdensity bin, all pairs of multiplicity functions that were possible to calculate in the bin were taken. Assuming Gaussian statistics, the probability that both multiplicity functions are compatible in each mass bin given their statistical uncertainties was estimated. The agreement between multiplicity functions was then quantified by computing the relative

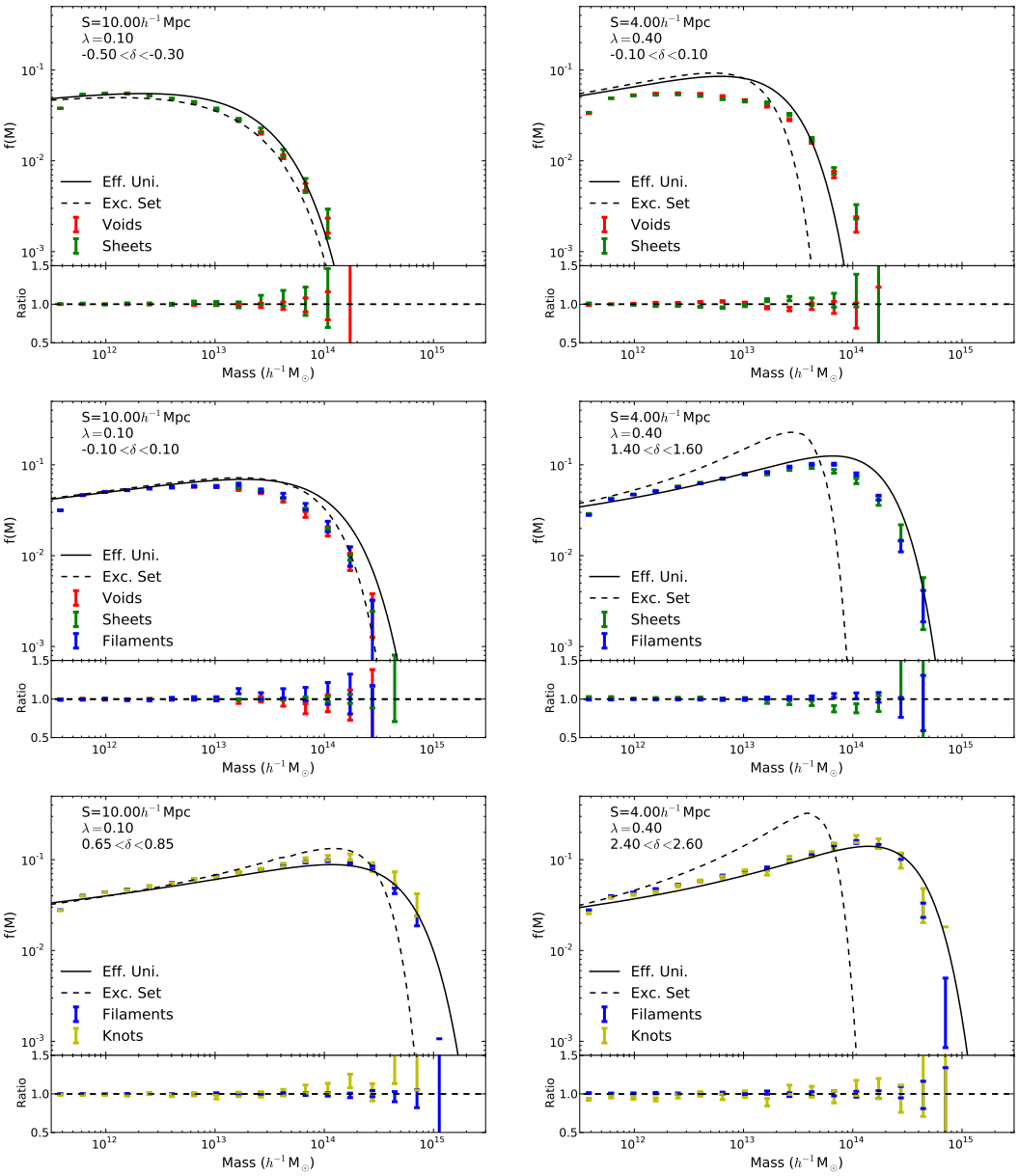


Figure 2.10 Multiplicity functions for the four different environments with their local densities restricted to a given range. Solid and dashed lines show the theoretical predictions of the effective-universe approach and the standard excursion set model respectively, data points show the multiplicity functions measured from the MDR1 simulation. In each plot, the lower panel shows the ratio between the different multiplicity functions and their mean value. The theoretical treatment predicts the same function for all environments in this case, which is realised to a good approximation in all cases. Neither the excursion set prediction nor the effective-universe approach agree quantitatively with the simulated data in all cases, but an overall better agreement is obtained for the effective-universe formalism, especially for smaller scales and larger environmental overdensities.

$S = 10\,h^{-1}\mathrm{Mpc},\ \lambda_{\rm th} = 0.1$		$S = 4\,h^{-1}\mathrm{Mpc},\ \lambda_{\rm th} = 0.4$	
δ	$\Delta f(M)\,(\%)$	δ	$\Delta f(M)\,(\%)$
$(-0.5, -0.3)$	1.1%	$(-0.1, 0.1)$	4.6%
$(-0.1, 0.1)$	13.2%	$(1.4, 1.6)$	3.5%
$(0.65, 0.85)$	5.2%	$(2.4, 2.6)$	8.6%

Table 2.1 Compatibility of the mass functions for different environments with restricted environmental densities for the two combinations of $(S, \lambda_{\rm th})$ explored in Figure 2.10: $(10\,h^{-1}\mathrm{Mpc}, 0.1)$, left column, and $(4\,h^{-1}\mathrm{Mpc}, 0.4)$, right column.

difference between them in the mass bin with the smallest p-value. The results are shown in Table 2.1 for the bins of overdensity explored above. In the worst case, the largest deviation is about 13%. However it is worth noting that, even in this case, both multiplicity functions are fully compatible, with a minimum p-value of 0.32, and in all other cases the values of the different multiplicity functions are compatible within 2-σ.

Even though the prediction that the halo abundances in different environments depend only on the environmental density has been verified, the exact dependence of these abundances on halo mass is not reproduced accurately by either the effective universe approach or the excursion set formalism, although they both qualitatively follow the same trend as the data.

2.5 Summary and discussion

This chapter has considered the statistics of dark-matter halos within the cosmic web, using the eigenvalues of the Hessian of the potential to classify regions of space into one of four geometrical environments. The main results of this chapter are:

1. Assuming the density contrast field to be Gaussian, clear predictions can be made regarding the abundance (i.e. volume and mass fractions) of the different environments classified according to the tidal tensor prescription. These are reasonable approximations for large smoothing scales and can be used to select eigenvalue thresholds that are useful for practical comparisons – partitioning the halo mass within the Universe nearly equally between the four environments.

2. The simulated halo abundances were compared with the predictions within the excursion set formalism and the effective-universe approach. Neither of these approaches are able to yield quantitatively precise results. However, the effective-universe picture provides an overall better description, especially for small smoothing scales and large environmental densities.

3. The Gaussian approach predicts that the only local property of the environment on which the conditional mass function depends is the density contrast δ_e. Thus a prediction of the mass function in the different geometrical environments should be possible if we know the overall dependence of the mass function on overdensity, and if we can predict the overdensity distributions for the different environments.

4. A detailed test of this prediction does not succeed very well, since the overdensity distributions are not well predicted by the Gaussian theory. This is improved in part by considering a lognormal model for the evolved density field, but discrepancies with numerical data remain.

5. Nevertheless, the fundamental prediction of this work has been directly tested, which is that different geometrical environments with the same overdensity should have identical halo mass functions. This is verified in the MDR1 simulation to a good approximation for a wide range of masses, filters and eigenvalue thresholds. A maximum relative deviation of about 13% between mass functions in different environments was found, which are, nevertheless, fully compatible in a statistical sense. In this regard, no evidence for any effect of tidal forces on halo abundances in addition to the impact of local overdensity has been seen. This could be consistent with the claim by Yan et al. (2013) that galaxy properties in the SDSS lacked an explicit dependence on environmental ellipticity, as well as with the results found by Metuki et al. (2015) in N-body simulations; it will be interesting to repeat such an analysis using the explicit decomposition by geometrical environment that has been studied here.

6. This result suggests that scalar halo properties are not heavily influenced by the tidal field beyond the local overdensity. This is not at odds with the results found by Forero-Romero et al. (2014) and Libeskind et al. (2014), who find that directional quantities, such as orientations or angular

momenta, show a strong correlation with the directions defined by the tidal tensor.

Whilst this chapter has focused on simulated data, with a successful extraction of the cosmic web from the MDR1 simulation, it is a logical next step to ask how such an analysis may be adapted to consider data from the real Universe as measured by observational surveys. It would be interesting to test and verify the lack of a strong influence of the tidal field on scalar halo properties within observational datasets. An extraction of the cosmic web from observational data also allows for various other test of the influence of the tidal field; can a similar statement be made regarding the scalar properties of observed galaxies? The following chapter details the application of the tidal tensor prescription to the observational data of the Galaxy And Mass Assembly spectroscopic redshift survey, drawing on the analyses presented in this chapter in order to investigate and quantify the additional uncertainties introduced due to the additional limitations of the dataset.

Chapter 3

The observed cosmic web

The previous chapter illustrated how and why one may attempt to classify the cosmic web within the output of large numerical simulations. However, in order to learn about the real Universe, it is a logical next step to extend the analysis to observational data. Part of the appeal of the tidal tensor prescription introduced in Chapter 2 is that it requires only information on the gravitational potential field, which can be derived from the underlying density field, and hence the method can be applied to both simulated and observational datasets. This chapter presents an application of the tidal tensor prescription to data from the Galaxy And Mass Assembly (GAMA) survey in order to extract the components of the observed cosmic web.

Numerous additional difficulties are encountered when working with observed rather than simulated data. For example, one no longer has a single time slice, of well defined and constant resolution, fully covering a large and periodic cubic volume. The dataset is limited, of variable accuracy and relies on observables such as galaxies rather than the underlying dark matter.

The GAMA survey and its galaxy, halo and mock catalogues are introduced in Section 3.1. Section 3.2 details thorough analyses, drawing on the previous work on simulated data, in order to quantify and reduce some of the uncertainties associated with the application of the tidal tensor prescription to observational data. Finally, Section 3.3 presents the resulting geometric environments defined within the GAMA survey and discusses how these compare to other studies of the cosmic web within GAMA.

Survey Region	RA range (J2000) (deg)	Dec. range (J2000) (deg)
G09	129.0 - 141.0	−2.0 - +3.0
G12	174.0 - 186.0	−3.0 - +2.0
G15	211.5 - 223.5	−2.0 - +3.0

Table 3.1 The GAMA equatorial survey regions.

3.1 GAMA: Galaxy And Mass Assembly

The GAMA survey (Driver et al. 2009, Driver et al. 2011, Liske et al. submitted) combines data from eight ground-based facilities and four space missions to survey the low-redshift galaxy population. Its main component is the spectroscopic survey of $\approx 300,000$ galaxies to $r < 19.8$ mag over ≈ 290 deg^2, split between five survey regions and conducted with the 2dF/AAOmega facility on the 3.9-m Anglo Australian Telescope (AAT). GAMA is an intermediate redshift survey, bridging the gap between wide-field surveys such as 2dFRGS (Colless et al. 2001) and SDSS (York et al. 2000) and high-redshift deep field surveys such as VIPERS (Garilli et al. 2014). By surveying a cosmologically representative volume whilst maintaining an impressive $> 98\%$ redshift completeness in the equatorial fields (Robotham et al. 2010), GAMA provides an ideal dataset with which to study the cosmic web and the modulation of galactic properties by large-scale environments.

3.1.1 The galaxy catalogue

After an expansion of the survey in 2010, the surveyed region was extended from three to five fields. This thesis uses data from the initial three equatorial fields, G09, G12 and G15 only, due to the high completeness in these regions. The equatorial fields each cover 12×5 deg^2 to an r-band Petrosian magnitude limit of 19.8 mag, over the coordinates given in Table 3.1.

3.1.2 Mock catalogues

A series of realistic mock galaxy catalogues were created by members of the GAMA team to enable tests and calibrations of analyses to be applied to the real data. The mock catalogues were constructed from the Millennium dark matter simulation (Springel et al. 2005) and populated with galaxies using the GALFORM (Bower et al. 2006) semi-analytic galaxy formation recipe. A

lightcone effect to mimic the observed redshift range was created by interpolating the galaxies between snapshots of the simulation. Due to the high spectroscopic completeness of GAMA, no incompleteness was incorporated into the mock catalogues. Similarly, velocity and magnitude uncertainties are not accounted for. In total 9 mock catalogues were extracted from the simulation with the exact survey geometry and angular separation of the three equatorial GAMA fields.

3.1.3 The group catalogue

The GAMA group catalogue is constructed from galaxy-galaxy linking via a Friends-of-Friends (FoF) algorithm, as described in detail in Robotham et al. (2011). The main feature of the algorithm is that it is anisotropic: the linking length is greater in the radial direction to allow for the 'finger of God' stretching due to the orbital motion within the groups. Extensive use was made of the mock catalogues in order to optimise the parameters of the FoF algorithm and test its robustness. The resulting groups extracted from the mock catalogue by the FoF algorithm were found to have median unbiased velocity dispersion and radius. 77% of the recovered FoF groups unambiguously matched a mock group and 89% of all mock groups were unambiguously recovered.

The estimation of group properties such as mass was also optimised using the mock catalogues. A dynamical group mass estimate was produced from the assumption that, for a virialized system, we expect its dynamical mass to scale as $M \propto \sigma^2 R$, where σ and R are unbiased estimates of the group velocity dispersion and group radius respectively. A scaling factor, A, was introduced to create a median unbiased estimate of $M_{\rm DM}/M_{\rm FoF}$, and determined in a semi-empirical manner by utilising the mock catalogues. Optimising over all mock groups with at least five members and requiring A to be constant results in $A = 10.0$, and the resultant dynamical mass estimate for each group is given by

$$\frac{M_{\rm FoF}}{h^{-1}{\rm M}_\odot} = \frac{A}{\frac{G}{{\rm M}_\odot{}^{-1}{\rm km}^2{\rm s}^{-2}{\rm Mpc}}} \left(\frac{\sigma}{\rm km\ s^{-1}}\right)^2 \frac{R}{h^{-1}{\rm Mpc}}. \tag{3.1}$$

The total group luminosities were estimated in a similar manner, introducing a scaling factor, B, which was again optimised using the mock catalogues. The total group luminosity was estimated by calculating the effective absolute magnitude

limit of each group and integrating over the global GAMA luminosity function in order to account for missing flux due to selection effects.

3.2 Classifying the cosmic web in observational datasets

In order to classify the cosmic web within GAMA, the tidal tensor prescription introduced in Section 2.1.2 is chosen once more for its applicability to both simulated and observational datasets and for its appealing theoretical underpinnings.

3.2.1 Measuring the tidal tensor

The tidal tensor is defined as the matrix of second derivatives of $\tilde{\Phi}$:

$$T_{ij} = \frac{\partial^2 \tilde{\Phi}}{\partial r_i r_j}, \tag{3.2}$$

where the usual definition of the Newtonian potential has been normalised to

$$\tilde{\Phi} \equiv \Phi/(4\pi G\bar{\rho}). \tag{3.3}$$

To calculate T_{ij} requires the matter overdensity field, δ. When working with observational data, for which direct knowledge of the underlying dark matter is not available, δ can be estimated from the galaxy overdensity field, $\delta_{\rm g}$, from which one can obtain a pseudo-gravitational potential satisfying $\nabla^2\tilde{\Phi_{\rm g}} \equiv \delta_{\rm g} = b\nabla^2\tilde{\Phi}$ in the limit of linear bias, b. The uncertainties introduced through using galaxies – biased tracers in redshift space – to estimate the real-space density field are discussed in Section 3.2.5. There an analysis of simulated data is shown, which indicates that using galaxies to estimate the underlying density field changes the classifications for <20% of the volume.

Galaxies are assigned to a Cartesian grid with cubic cells of width $R_c = 3\,h^{-1}$Mpc by cloud-in-cell interpolation. Experimentation with the value of R_c in simulated data, as shown in Fig. 2.1 of the previous chapter, showed that results converge by $R_c \approx 4\,h^{-1}$Mpc, for density fields smoothed with a Gaussian filter of effective width $\sigma \geq 4\,h^{-1}$Mpc, and any further variation caused by using smaller grid cells is negligible. A standard ΛCDM cosmology with $\Omega_{\rm M} = 0.25$, $\Omega_\Lambda =$

0.75 and $H_0 = 100\,h\,\mathrm{kms}^{-1}\mathrm{Mpc}^{-1}$ is assumed for the conversion to Cartesian coordinates. Note that apart from during this gridding process, the classification of geometric environments implemented in this work is cosmology independent.

The overdensity of each cell is given by

$$\delta_\mathrm{g} = \frac{N_\mathrm{obs}}{N_\mathrm{R}} - 1, \tag{3.4}$$

where N_obs is the number of observed galaxies within the cell after the interpolation, and N_R is an estimate of the corresponding number which would have been observed if there were no clustering. More specifically, nN_R is the interpolated number density of a randoms catalogue generated by cloning real GAMA galaxies in the sample n times ($n = 400$ is used here) and distributing them randomly within the maximum volume over which they can be observed (Farrow et al. in prep., Cole 2011). As with the analysis of simulated data, to suppress shot noise, and also to remove extreme non-linearities, the density field is smoothed with a Gaussian filter of variable width, σ_s. Again, the smoothing due to the interpolation is taken into account, leading to an effective smoothing scale of $\sigma = \sqrt{(\sigma_\mathrm{s}^2 + R_\mathrm{c}^2/6)}$. In the spirit of the Zel'dovich approximation, an appropriate filter should reach scales where only a moderate amount of shell-crossing has occurred, linking the observed density field to the initial conditions. With this, and the number density and survey geometry of GAMA in mind, the effective smoothing scales of $\sigma = 4$ and $10\,h^{-1}\mathrm{Mpc}$ are chosen. Note that these are the same smoothing scales as used in the analysis of simulated data in Chapter 2.

An immediate practical problem is how to deal with the survey boundaries during this smoothing process given that there is no information on the density field beyond the surveyed region. Zero-padding the survey, by setting $\delta_\mathrm{g} = 0$ for regions outside of the survey boundaries, would bias the density field inside the survey. In order to ameliorate this problem, before the smoothing process the volume outside of the survey is populated with cloned galaxies 'reflected' along the boundaries of the field, which is approximately equivalent to using a weighted smoothing kernel. This method of reflecting cloned galaxies is discussed in more detail in Section 3.2.4.

The pseudo-gravitational potential field and its second spatial derivatives can be derived from the smoothed galaxy-overdensity field by working in Fourier

space. The potential, $\Phi_{\rm g}$, can be obtained by solving Poisson's equation

$$\nabla^2\Phi_{\rm g} = 4\pi G\bar{\rho}\delta_{\rm g} = \alpha + \beta + \gamma, \tag{3.5}$$

where $\bar{\rho}$ is the average matter density of the Universe, G the gravitational constant and α, β, γ are the 3 eigenvalues of the diagonalised Hessian of $\Phi_{\rm g}$. However, here we consider the dimensionless potential, $\tilde{\Phi}_{\rm g}$, and the dimensionless eigenvalues λ_1, λ_2 and λ_3, found by factoring $4\pi G\bar{\rho}$ out of Eq. 3.5:

$$\nabla^2\tilde{\Phi}_{\rm g} = \delta_{\rm g} = \lambda_1 + \lambda_2 + \lambda_3. \tag{3.6}$$

Note that in the limit of linear bias, any bias can be absorbed by a rescaling of the eigenvalue threshold.

In Fourier space the dimensionless potential and its Hessian, the tidal tensor, T_{ij}, are given by

$$\tilde{\Phi}_k = -\frac{\delta_k}{k^2} \quad \text{and} \quad T_{ij} = \frac{\partial^2\tilde{\Phi}_k}{\partial_i\partial_j} = \frac{k_i k_j \delta_k}{k^2}, \tag{3.7}$$

with $k = \sqrt{(k_i^2 + k_j^2 + k_k^2)}$.

The eigenvalues of T_{ij} are calculated for each cell of the Cartesian grid and comparison with an eigenvalue threshold, discussed in Section 3.2.3, leads to the classification of the region within the cell.

3.2.2 Sample selection

A significant and problematic difference between observational and simulated datasets is that, for unavoidable practical reasons, the selection effects vary with redshift in observational data. Whilst bright galaxies may be observed over the full redshift range of a survey, fainter galaxies may only be bright enough to be observed at low redshift. Hence, the type of galaxies included in observational catalogues may vary significantly across a large redshift range. GAMA observes to $r < 19.8$ mag and $z \lesssim 0.5$, however, it is important to test whether any classification of geometric environments is robust over the full survey or if a limited subset of the data is more optimal.

In addition to the selection effects, as this analysis is derived from the galaxy

overdensity field over a large redshift range, the variation of the galaxy bias and of the growth factor with redshift must be considered. The galaxy bias, $b(z)$, describes how the galaxy overdensity field, δ_g, is related to the matter overdensity field, δ_m, by

$$\delta_g = b(z)\delta_m. \tag{3.8}$$

Cosmic time and the expansion of the Universe both have an impact on the growth of structures; in the linear regime, the evolution of matter fluctuations, $\delta_{\rm m}(\mathbf{r}, z)$, with redshift is described by

$$\delta_{\rm m}(\mathbf{r}, z) = D(z)\delta_{\rm m}(\mathbf{r}, z_0), \tag{3.9}$$

where $D(z)$ is the growth factor and z_0 is an arbitrarily chosen reference redshift. Ideally the classification of geometric environments should be redshift independent. For example, a region classed as a filament at $z = 0.5$ should also be identified as a filament if it were to be observed today, if the geometric environments are defined by initial conditions. Any modification of the definitions of environments that are caused by effectively allowing the eigenvalue threshold to vary with redshift due redshift dependent biases should be eliminated as much as possible. Using the galaxy bias and growth factor, Eq. 3.6 may be extended to

$$\delta_{\rm g}(z) = \bar{b}(z)D(z)\delta_{\rm m}(z_0) = \lambda_1 + \lambda_2 + \lambda_3, \tag{3.10}$$

and we may consider fluctuations in the galaxy overdensity field to follow

$$\delta_g(\mathbf{r}, z) = \delta_g(\mathbf{r}, z_0)\frac{\bar{b}(z)D(z)}{\bar{b}(z_0)D(z_0)} \equiv \delta_g(\mathbf{r}, z_0)A(z). \tag{3.11}$$

It follows that

$$\sum_{i=1}^{3} \lambda_i(z) = A(z)\sum_{i=1}^{3} \lambda_i(z_0) \tag{3.12}$$

and, hence, for whichever threshold is chosen, one may argue that it should scale with redshift as $\lambda_{\rm th}(z) = A(z)\lambda_{\rm th}(z_0)$.

The value of $A(z)$ can be estimated theoretically by modelling the bias and the growth factor as a function of redshift separately. The bias of observed galaxies varies with redshift due to the varying range of luminosities which fall within the survey magnitude limit. If the minimum luminosity required for a galaxy to be

observable at redshift z is $L_{min}(z)$, then the average bias at redshift z is found by integrating the luminosity function over all observable luminosities,

$$\bar{b}(z) = \frac{\int_{Lmin(z)}^{\infty} b(L)n(L)dL}{\int_{Lmin(z)}^{\infty} n(L)dL}, \tag{3.13}$$

where $n(L)$ is the number of galaxies per unit luminosity per unit volume and $b(L)$ is the estimated bias of a galaxy with luminosity L. By converting redshift to luminosity distance, $d(z)$, the minimum luminosity is given by

$$L_{min}(z)/L_{\odot} = 10^{(\frac{m_{\odot}-m_{r,max}}{2.5})} \left(\frac{d(z)}{d\odot}\right)^2, \tag{3.14}$$

where $m_{r,max} = 19.8$ is the GAMA magnitude limit. To estimate the bias as a function of luminosity, the results of Norberg et al. (2001) are used. They find that $\bar{b}(L)$ is well fit by

$$\bar{b}(L) = a_0 + a_1 \left(\frac{L}{L^*}\right), \tag{3.15}$$

with $(a_0, a_1) = (0.85, 0.15)$, and L^* is a characteristic luminosity found from a Schechter function fit to the luminosity function, given by

$$n(L) = \phi(L) = \left(\frac{\phi^*}{L^*}\right) \left(\frac{L}{L^*}\right)^{\alpha} e^{-L/L^*}. \tag{3.16}$$

The resulting average bias as a function of redshift is shown in Fig. 3.1. It can be seen that the bias increases smoothly with redshift, and approximately doubles over the full redshift range of the GAMA survey.

To model the growth factor, $D(z)$, the analytic approximation discussed in Hoffmann et al. (2015) is used:

$$D(z) \propto \frac{(1+z)^2}{H^2(z)} \left[f(z) + 1 + \frac{\Omega_m(z)}{2} - \Omega_{\lambda}(z)\right]^{-1}, \tag{3.17}$$

where, for a spatially flat Universe, $f(z) \equiv [\Omega_m]^{\alpha(z)}$, and

$$\alpha(z) \simeq \frac{6}{11} - \frac{15}{2057} \ln[\Omega_m(z)] + \frac{205}{540421} \ln^2[\Omega_m(z)]. \tag{3.18}$$

When normalised to 1 at $z = 0$, the resulting growth factor as a function of redshift is shown in Fig. 3.2. It can be seen that, contrary to $\bar{b}(z)$, the growth

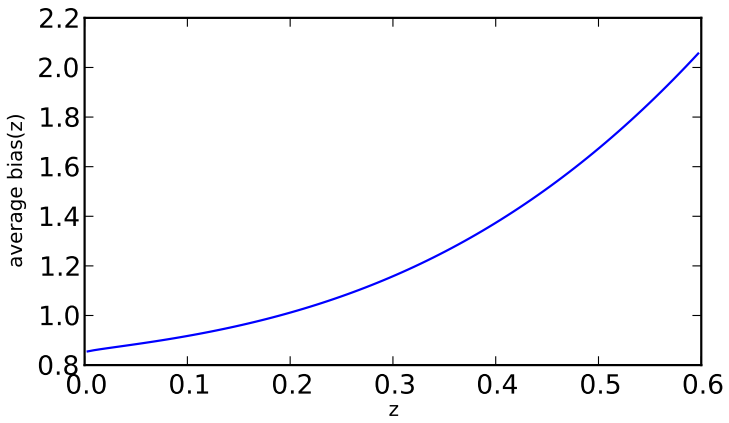


Figure 3.1 Estimated average bias, $\bar{b}(z)$, of galaxies measured within the GAMA survey as a function of redshift

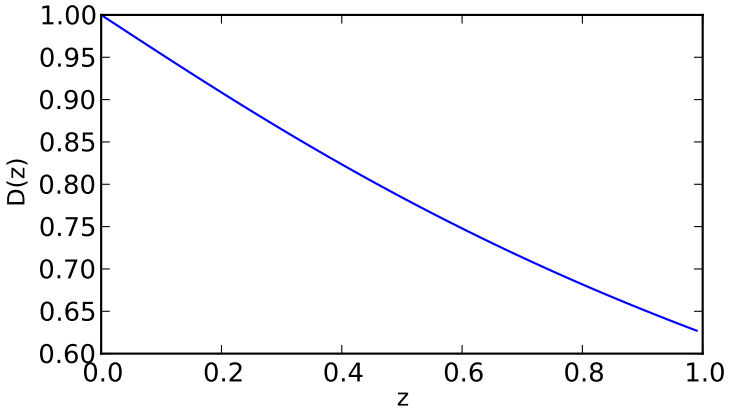


Figure 3.2 Growth factor, $D(z)$, estimated from Eq. 3.17, as a function of redshift and normalised to $D(z = 0) = 1$

factor is an approximately linear *decreasing* function of redshift. To some extent, then, one might hope that the two opposing factors cancel out, reducing the redshift dependence of the resulting classifications.

The multiplication of the redshift dependent growth factor and bias, when both are normalised by their values at $z = 0$, gives our estimate for $A(z)$, the required scaling factor for a redshift dependent threshold. This combination is shown in Fig. 3.3, and can be seen to be relatively flat at low redshift and increase substantially beyond $z \approx 0.25$. The red line of the figure shows a non-linear least squares fit to the resulting estimate of $A(z)$, as given in the key.

The necessity and benefit of imposing such a redshift dependent threshold

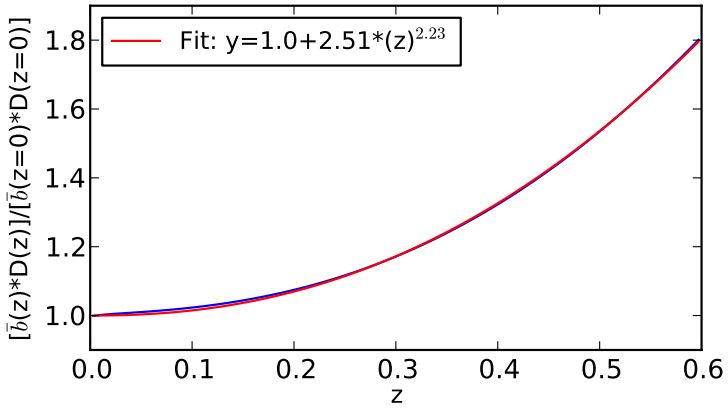


Figure 3.3 Blue: The estimated average bias multiplied by the estimated growth factor, normalised by their values at $z = 0$, as a function of redshift. **Red:** A non-linear least squares fit, as given in the key.

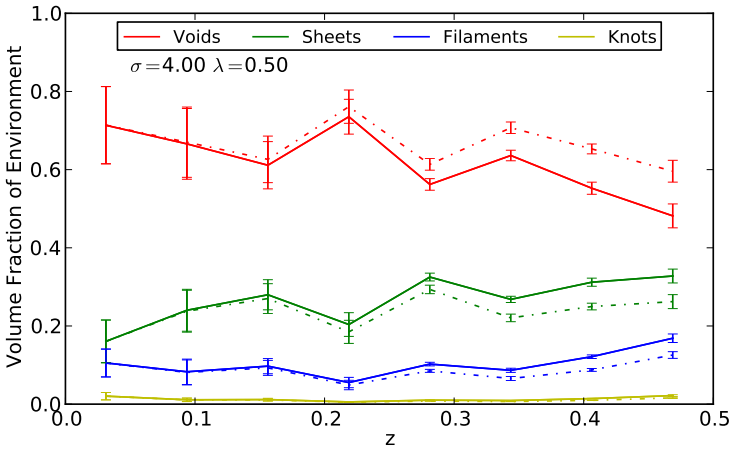


Figure 3.4 Fraction of volume within redshift bins classified as each environment, after smoothing with Gaussian filter of width $\sigma = 4\,h^{-1}\mathrm{Mpc}$, averaged over the three GAMA fields. Error are standard errors on the mean of the values for each field. **Solid lines:** A constant threshold of $\lambda_{\mathrm{th}} = 0.5$. **Dot-Dashed Lines:** A redshift dependent threshold, $\lambda_{\mathrm{th}}(z) = A(z)0.5$, scaled by theoretical estimation of $A(z) = b(\bar{z})D(z)$, as described in the text.

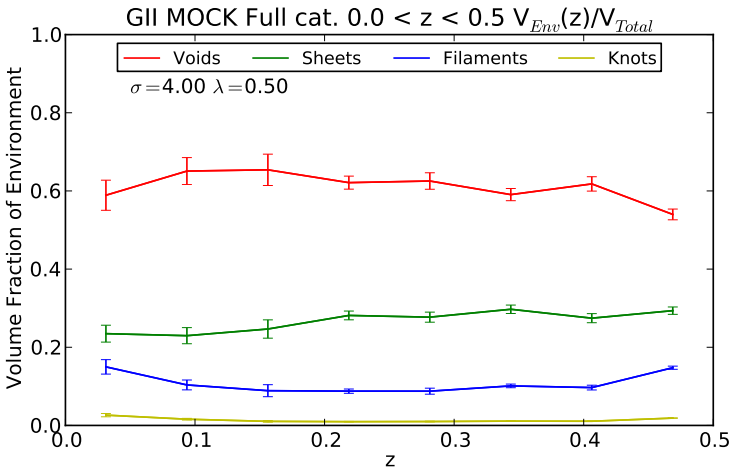


Figure 3.5 Fraction of volume within redshift bins classified as each environment. Averaged over the nine realisations of the GAMA mocks. Errors are standard errors on the mean of the values for each field. $\sigma = 4\,h^{-1}\mathrm{Mpc}, \lambda_{\mathrm{th}} = 0.5$.

was investigated by measuring how the fraction of the volume classified as each environment varies within redshift bins. Fig. 3.4 shows these volume fractions within redshift bins of width $\Delta z = 0.05$, as measured from the three GAMA fields, with error bars to represent the standard error on the mean within each bin. The dot-dashed line shows the equivalent fractions found when the eigenvalue threshold is modified as discussed, by the power law fit to the scaling factor $A(z)$ given in the key of Fig. 3.3. It can be seen that this redshift dependent threshold has little effect on the environments at low redshift, and causes the fractions at higher redshift to be slightly more constant with redshift. It should also be noted that, due to the smaller volume of the survey available at low redshift, the error bars are more significant, allowing for larger expected deviations from a constant volume fraction. For this reason, the investigation was repeated on the GAMA mocks, calculating the average volume fraction of each environment over the 9 realisations, as shown in Fig. 3.5. The 9 realisations provide more statistics, reducing the uncertainty on the expected result, and a significantly more stable classification of geometric environments across the full redshift range is seen.

Whilst there may be arguments in favour of imposing such a redshift dependent threshold, the additional complications, uncertainty in the correct value of $A(z)$, and relatively constant abundance of environments seen in Fig. 3.5 when analysing the mock data add to appeal of the simpler, constant threshold.

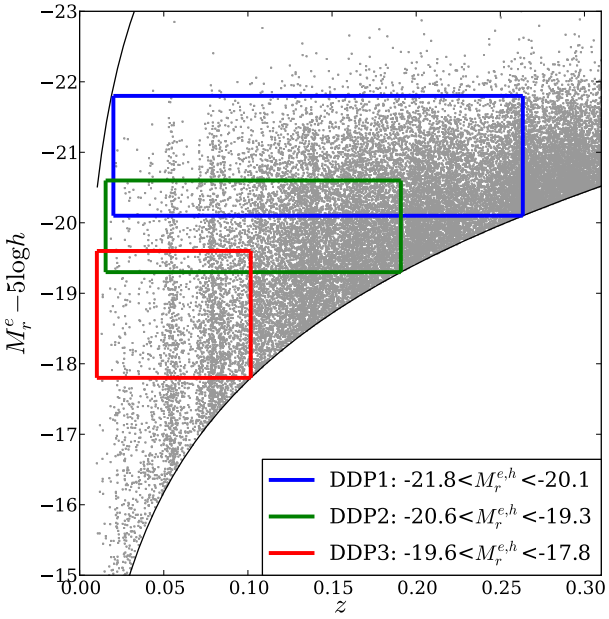


Figure 3.6 The three volume-limited samples considered in McNaught-Roberts et al. (2014). The blue box illustrates the galaxies selected for the volume-limited sample tested in this work.

For this reason, henceforth, an eigenvalue threshold which is constant across all redshifts is used. However, due to large increase of $A(z)$ at high redshift seen in Fig. 3.3, the focus is placed on the lower redshift regime, where the variation in the bias and growth factor is expected to be less significant. The redshift cuts used in previous GAMA work (McNaught-Roberts et al. 2014) which focus on the low redshift portion of the survey, namely, $0.04 < z < 0.263$, are applied throughout the remaining analyses of the GAMA data. An extension to the full redshift range of the survey is left for future work.

With the redshift range set, the issue of redshift dependent selection effects should now be considered. The full GAMA catalogue is a magnitude limited sample, where all objects within the magnitude limit of the survey, $r < 19.8$, are included. A volume limited sample can be extracted from this, with absolute magnitude and redshift limits chosen such that any object within the sample could in principle have been observed across the full redshift range. The disadvantage of this is that it requires discarding many faint objects which could only be found at low redshift. For example, Fig. 3.6 shows three volume limited samples considered in McNaught-Roberts et al. (2014). All galaxies falling outside of the

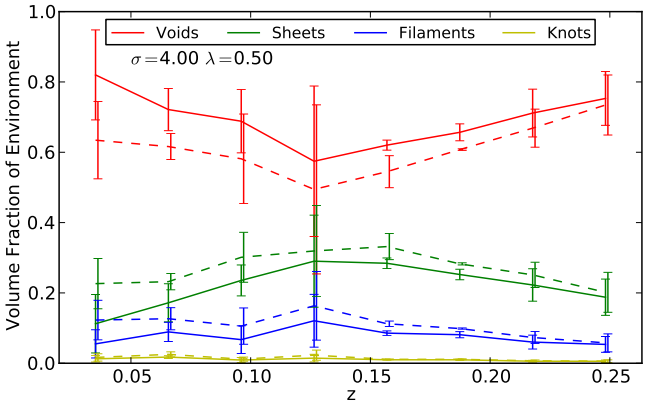


Figure 3.7 Fraction of volume within redshift bins classified as each environment. Averaged over the three GAMA fields. Errors are standard errors on the mean of the values for each field. $\sigma = 4\,h^{-1}\mathrm{Mpc}, \lambda_{\mathrm{th}} = 0.5$. **Dashed lines:** Volume limited sample. **Solid lines:** Magnitude limited sample.

redshift-magnitude limits depicted by the coloured boxes are discarded from the samples. The advantage of such volume-limited samples is that they sample a constant range of absolute magnitudes. The impact of the choice of a volume or magnitude limited sample on the resulting environment classifications was tested, again by calculating the volume fractions of each geometric environment across redshift bins to investigate the variation of these fractions with redshift. Fig. 3.7 displays the resulting fractions for all GAMA galaxies within the chosen redshift limits of $0.04 < z < 0.263$, and for the volume limited sample which can be extracted from that redshift range by imposing the absolute magnitude criteria, $-21.8 < M_r < -20.1$, as shown by the blue box in Fig. 3.6. Similarly, the volume fractions as a function of the eigenvalue threshold for the two samples are shown in Fig. 3.8. Fig. 3.7 indicates that the choice of sample has a non-negligible effect on the environment classifications, however, the redshift variations are still present. Fig. 3.8 indicates that selecting a volume limited sample has a similar effect to slightly reducing the eigenvalue threshold. With no strong evidence in favour of either sample, in order to avoid discarding objects unnecessarily, the magnitude limited sample is used and all galaxies within the chosen redshift limits are retained.

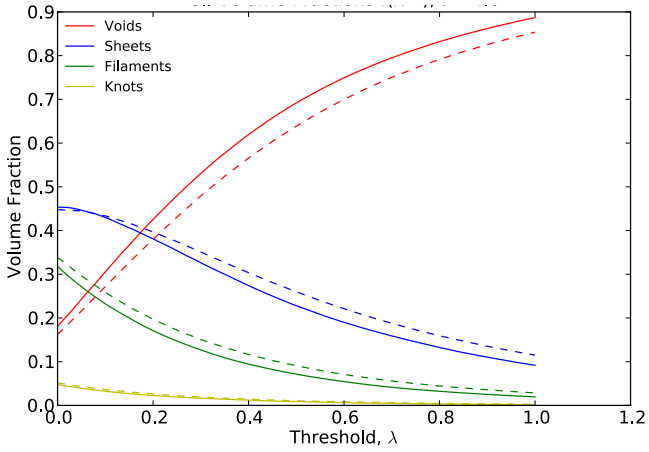


Figure 3.8 Fraction of volume classified as each environment as a function of the eigenvalue threshold, $\lambda_{\rm th}$. Averaged over the three GAMA fields. $\sigma = 4\,h^{-1}$Mpc. **Dashed lines:** Volume limited sample. **Solid lines**: Magnitude limited sample.

3.2.3 The eigenvalue threshold

To classify the geometric environments within this GAMA sample, one must choose the value at which to set the eigenvalue threshold, $\lambda_{\rm th}$. The number of eigenvalues greater than this threshold is used to define environments rather than the number greater than zero. After the normalisation discussed in Section 3.2.1, Eq. 3.6 shows that the sum of the eigenvalues will be equal to the galaxy density contrast, hence an appropriate threshold parameter will be of order unity.

With the introduction of $\lambda_{\rm th}$, the three eigenvalues calculated for each location leads to the classification of regions in the same way as for dark matter in Chapter 2 (with $\lambda_3 < \lambda_2 < \lambda_1$):

- **Voids**: all eigenvalues below the threshold ($\lambda_1 < \lambda_{\rm th}$).
- **Sheets**: one eigenvalue above the threshold ($\lambda_1 > \lambda_{\rm th}$, $\lambda_2 < \lambda_{\rm th}$).
- **Filaments**: two eigenvalues above the threshold ($\lambda_2 > \lambda_{\rm th}$, $\lambda_3 < \lambda_{\rm th}$).
- **Knots**: all eigenvalues above the threshold ($\lambda_3 > \lambda_{\rm th}$).

This work presents results for two smoothing scales, $\sigma = 4$ and $10\,h^{-1}$Mpc, chosen to study a wide range of scales whilst reflecting the limitations caused by the number density and survey volume of GAMA. The choice of $\lambda_{\rm th}$ is similarly arbitrary; whilst it changes the classification of the web, there is no strict definition of what constitutes a void region for example, and hence the classifications can be adapted to suit the task at hand. One could use the spherical collapse model to explicitly derive the eigenvalue threshold that corresponds to collapse along the eigenvector by equating the collapse time with the age of the Universe, but the invalid assumption of spherical isotropic collapse would allow for only a rough estimate of the true threshold. An alternative approach is to choose the threshold that produces the best visual agreement of the resulting web with the distribution of matter, however, this choice is rather subjective and not easy to justify scientifically. Instead, in this work, the value of the eigenvalue threshold is chosen in order to optimise the statistical significance of any measurement that may be made in the different environments, i.e. to allocate the objects under study to the four environments as equally as possible. To do so, for a variety of parameter sets, the root mean square dispersion (RMSD) of the fraction of all galaxies in the selected sample classified as each of the four geometric environments from the mean fraction was calculated (cf. Eq. 2.4 in Chapter 2). Fig. 3.9 shows this root mean squared dispersion in environmental number count as a function of the smoothing scale, σ, and the imposed eigenvalue threshold, $\lambda_{\rm th}$. This figure is equivalent to Fig. 2.4 of Chapter 2, except that here the partitioning of galaxies rather than dark matter is calculated and again note that the bias alters the meaning of $\lambda_{\rm th}$ when the pseudo-potential derived from the galaxy density field is used. Nonetheless, the two figures are qualitatively similar and both show a comparable optimal parameter space.

This RMSD should be minimised in order to ensure that all environments hold enough galaxies to maintain a low level of statistical uncertainty, which is essential in order to look for potentially small modulations due to geometric environments. No choice of parameters produces an exactly equal split, where each environment holds 25% of galaxies, but there are a range of parameters such that each environment holds at least 10%. The dark shaded region represents this optimal parameter space – for smaller smoothing scales a higher threshold is required in order to maintain a near-comparable split of galaxies and vice versa.

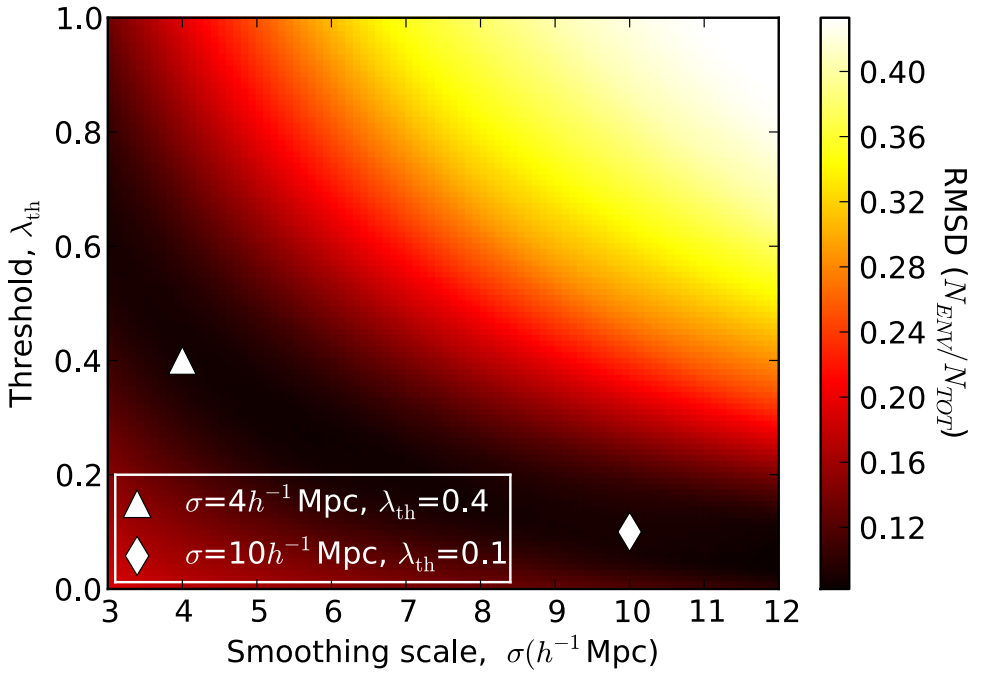


Figure 3.9 The two free parameters, the eigenvalue threshold, $\lambda_{\rm th}$, and the smoothing scale, σ, are chosen in a way which optimises the resulting statistics by assigning a comparable number of objects to each geometric environment. This plot displays the root mean squared dispersion (RMSD, as defined by Eq. 2.4), in the number of galaxies in the sample which are assigned to each geometric environment as a function of $\lambda_{\rm th}$ and σ used to generate the classifications. The dark curve represents the statistically optimal region in the parameter space, motivating the choice of two parameter sets: (σ, $\lambda_{\rm th}$) = ($4\,h^{-1}{\rm Mpc}$, 0.4) and ($10\,h^{-1}{\rm Mpc}$, 0.1), as indicated in the figure.

Environment	$\sigma(\,h^{-1}\mathrm{Mpc})$	Volume (%)	Galaxies (%)	$\delta_8^{\bar{e}nv}$
Voids	4	59	18	−0.16
Sheets	4	29	34	0.81
Filaments	4	10	36	2.38
Knots	4	1	12	4.39
Voids	10	37	15	−0.03
Sheets	10	39	32	0.69
Filaments	10	20	39	1.93
Knots	10	3	15	3.82

Table 3.2 The percentage of the volume and the percentage of galaxies within each geometric environment, classified with either $(\sigma, \lambda_{\rm th}) = (4\,h^{-1}\mathrm{Mpc}, 0.4)$ or $(10\,h^{-1}\mathrm{Mpc}, 0.1)$. The final column gives the average local overdensity, $\delta_8^{\bar{e}nv}$, of galaxies in each environment, derived from the number of galaxies within a sphere of radius $8h^{-1}$Mpc.

Based on this, this work focuses on environments defined by the parameter sets shown by the symbols in the figure: $(\sigma,\lambda_{\rm th}) = (4\,h^{-1}\mathrm{Mpc},\, 0.4)$ and $(10\,h^{-1}\mathrm{Mpc},\, 0.1)$, the same parameters chosen for the study of simulated data in Chapter 2. The resulting partition of galaxies and of the survey volume between the environments defined by these two parameter sets are given in Table 3.2. Note that these parameters also produce environments with good visual agreement to the distribution of matter, as can be seen in Section 3.3.1.

3.2.4 Effects of the survey geometry

A significant limitation of the tidal tensor prescription, or of any analysis requiring information on the smoothed density field, is that observational datasets lack information beyond the surveyed region. There is then the question of how to smooth the galaxy distribution near the survey boundaries. If one 'zero-pads' the volume outside of the survey, by setting it all to the large-scale average overdensity ($\bar{\delta} = 0$ by definition) this will, on average, reduce the magnitude of both overdensities and underdensities that may be straddling the border of the survey and alter the resulting estimate of the true underlying density in a systematic but unpredictable way. To mitigate this effect, in this work the galaxies are 'reflected' along each field's boundaries in right ascension (ra) and declination (dec). The galaxies inside the survey volume are cloned and given an appropriate reflected location outside of the survey volume before the smoothing

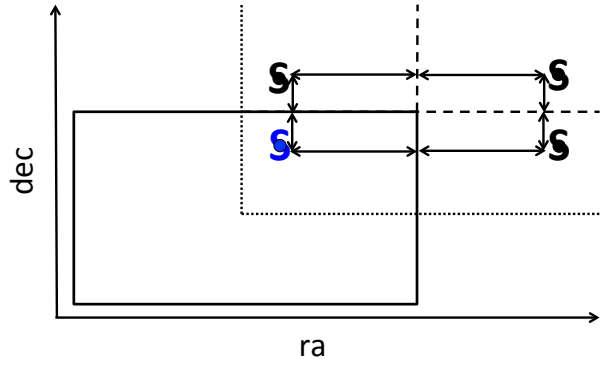
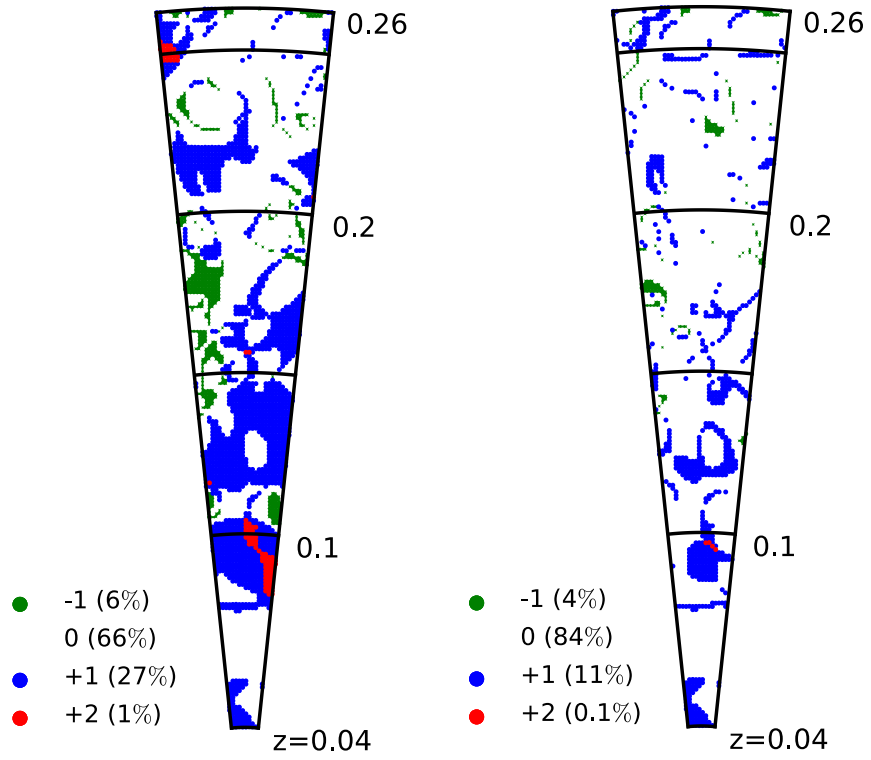


Figure 3.10 A sketch of the reflection method. Each galaxy in the survey sample, represented by the blue spiral, is cloned 3 times as depicted by the black spirals. The solid line box represents the ra and dec boundaries of the GAMA field, with redshift pointing out of the page. The dotted lines show the quadrant which is reflected along the nearest two boundaries of the field, which are of constant ra or dec and shown by the dashed lines. Each long arrow in the figure represents the same difference in right ascension, similarly each short arrow represents the same difference in declination.

process. In effect, the reasonable assumption that large-scale features continue smoothly beyond the survey edge is made.

A sketch illustrating the reflection process is shown in Fig. 3.10. Each galaxy is cloned 3 times always keeping its original redshift, the first clone is given a new ra equivalent to a reflection along the nearest ra border of the field, the second clone keeps the ra of the original galaxy but has its dec changed to simulate a reflection along the nearest dec boundary of the field, and the third takes on both of these two new ra and dec values. This has an approximately equivalent effect to using a weighted smoothing kernel, where cells near the edge of the volume are up-weighted to account for the lack of information in cells outside of the volume. Beyond the reflection regions (half the width of the survey dimension) zero-padding is used. In the redshift direction, for each field, the full GAMA galaxy catalogue, $0.0 < z < 0.5$ is used, and a density contrast for the full surveyed volume of each field is calculated, though, as described in the text, only galaxies satisfying $0.04 < z < 0.263$ are selected for the scientific analyses. The MultiDark $(1\,h^{-1}\mathrm{Gpc})^3$ dark matter simulation (Prada et al. 2012) discussed in Chapter 2 is used to test this reflection method, selecting GAMA-representative regions from the full simulation where necessary. The simulated dataset used is a single redshift snapshot of $z = 0.1$ with galaxies randomly sampled so that the



(a) Zero-padding **(b)** Reflected galaxies

Figure 3.11 A test of the effect of the survey geometry on the resulting geometric environment classifications using simulated data. Coloured regions of the figure show the cells that are classified differently to the full simulation results when regions outside a GAMA sized survey cone are zero-padded (**left** panel) or filled with reflected galaxies (**right** panel), as described in the text. This is for an example realisation of a GAMA field and (σ, $\lambda_{\rm th}$)=(10 h^{-1}Mpc, 0.1). Colour code in the keys refer to the difference, ΔN, in the number of eigenvalues above the threshold, N^{+}, between the full simulation and the limited-information survey classifications, e.g. $\Delta N = N^{+}_{\rm FULL} - N^{+}_{\rm 0pad}$. Hence each cell has a discrete value of ΔN, with $-3 \leq \Delta N \leq 3$. The percentage of all cells with a given ΔN, measured over three realisations, is indicated in the keys. A high percentage with $\Delta N = 0$ is desired, as this indicates the limited information has not changed the resultant environment classification of these cells. The increase in $\Delta N = 0$ from 66% to 84% shows that the reflection technique offers a strong improvement over zero-padding alone.

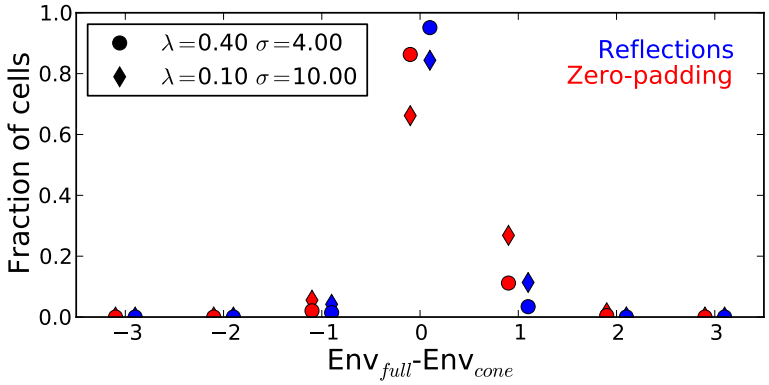


Figure 3.12 The cells that are classified differently, with **Diamonds:** (σ, λ)=(10 h^{-1}Mpc, 0.1), **circles:** (σ, λ)=(4 h^{-1}Mpc, 0.4), when regions outside a GAMA sized survey cone are **Red:** zero-padded and **Blue:** filled with reflected galaxies, as described in the text, and then zero padded. x-axis represents the difference in the number of eigenvalues above the threshold between the full simulation and the limited-information survey-style classification, averaged over three realisations of a GAMA field (with a small offset from the discrete values for clarity). The reflections are shown to reduce the effect of the limited information for both parameter sets.

number density of mock galaxies matches the average number density of galaxies in the GAMA sample. Fig. 3.11 shows, for the simulated data, the regions in which the environments resulting from the larger smoothing scale parameter set differ from those when information of the full periodic cube is used when only the GAMA-sized volume information is kept, and other regions are either zero-padded only, or populated with the cloned galaxies as described above. The quantitative results for both parameter sets are contrasted in Fig. 3.12; note that the 4 h^{-1}Mpc smoothing is less affected by the survey geometry (due a reduced 'skin-depth' of volume which is significantly affected by the volume outside of the survey), but shows a similar improvement when using this reflection technique. The differences are not confined to the edges of the survey due to the use of Fourier transforms but instead tend to occur along boundaries between regions of different environments due to a slight change in the calculated eigenvalues. The percentages of cells classified differently, measured over three realisations, are given in Fig. 3.12 and in the key of Fig. 3.11. It can be seen that the reflection technique is beneficial, increasing the percentage of correctly classified cells from 66% to 84% for the larger smoothing scale, so that the classifications more closely mimic the results of the full simulation than when zero-padding

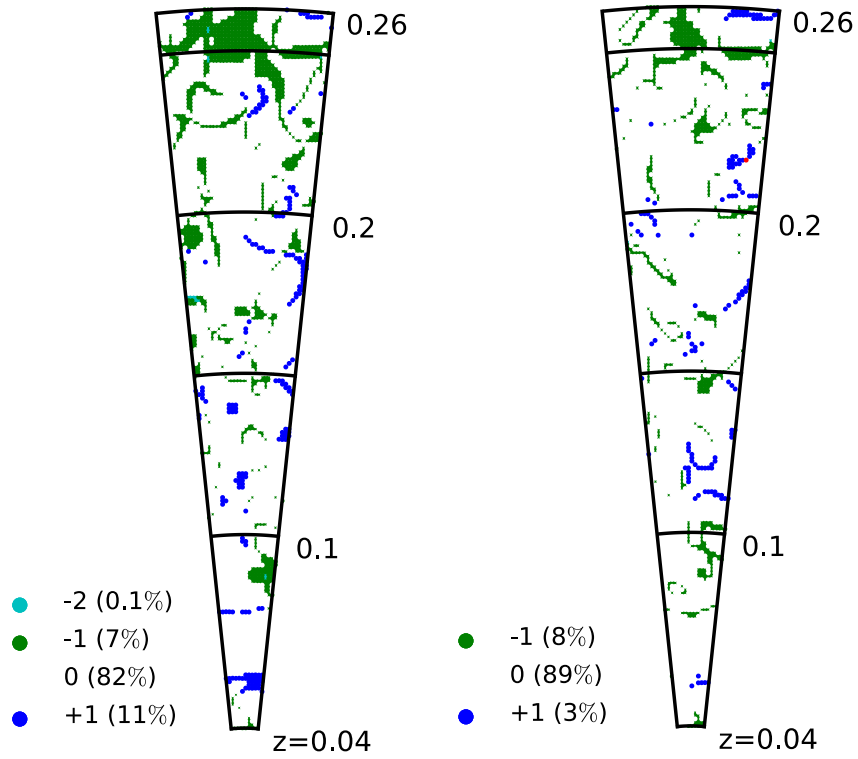
alone is used. There are remaining unavoidable discrepancies due to the lack of information, however, note that 99.9% of cells are classified within ± 1 dimension of environment from the 'full-information' results when the reflection technique is applied, an increase from the corresponding value for zero-padding of 99%.

3.2.5 Redshift-space distortions and galaxy bias

As discussed in section 3.2.1, the use of biased galaxies in redshift space to estimate the underlying matter overdensity requires caution. Again the simulated dataset is utilised to investigate the magnitude of these effects on the resulting environment classifications. The MDR1 simulation provides information on both the underlying dark matter density field and a simulated galaxy catalogue with galaxy velocity information. This enables a comparison of the resulting classifications when the density field is estimated from the locations of galaxies and when the underlying dark matter density field is used directly. In a similar manner to Fig. 3.11, Fig. 3.13a shows those cells that change their classification when the galaxy density field rather than the dark matter density field is used. The use of galaxies to estimate the density results in 20% of the volume appearing to belong to a different geometric environment. Results for both parameter sets are shown in Fig. 3.14, which indicates that environments calculated from both parameter sets experience a similar magnitude of uncertainty due to this density field estimation.

With the velocity information it is possible to shift each galaxy in the simulation to its redshift-space coordinates, by estimating the distance which would have been inferred given its location and radial velocity, and again compute this comparison. Fig. 3.13b and Fig. 3.14 compare the classifications for density fields constructed from redshift- and real-space galaxies. The redshift-space distortions are found to have to have no effect on $\sim 90\%$ of the volume for both 4 and $10\,h^{-1}$Mpc smoothing scales.

The combined effect of the three main causes of error when applying the tidal tensor prescription to observational data (survey geometry, a density field sourced from the galaxy number density and redshift-space distortions) is found to be a change in the resulting geometric environment of $< 25\%$ of the volume for both 4 and $10\,h^{-1}$Mpc smoothing scales. A example realisation of a field is shown in Fig. 3.15, indicating the regions which are classified differently when the three



(a) DM vs galaxy δ **(b)** Real- vs Redshift-space galaxies

Figure 3.13 A test of the effects of redshift-space distortions and of using a density field estimated from galaxy number counts on the resulting geometric environment classifications within simulated data. Coloured regions of the figure show the cells that are classified differently, with $(\sigma,\ \lambda_{\rm th})$=$(10\,h^{-1}{\rm Mpc},\ 0.1)$, when the dark matter density field is used or the density field is calculated from the (real-space) galaxies (**left** panel) and when the density field is calculated from the real-space galaxies or from redshift-space galaxies (**right** panel). Colour code in the keys refer to the difference in the number of eigenvalues above the threshold, N^{+}, between the full simulation and the limited-information survey-style classifications, e.g. $N^{+}_{\rm DM} - N^{+}_{\rm gal}$ or $N^{+}_{\rm real-sp} - N^{+}_{\rm redshift-sp}$. The percentages of cells with each difference value, measured over three realisations, are indicated in the keys.

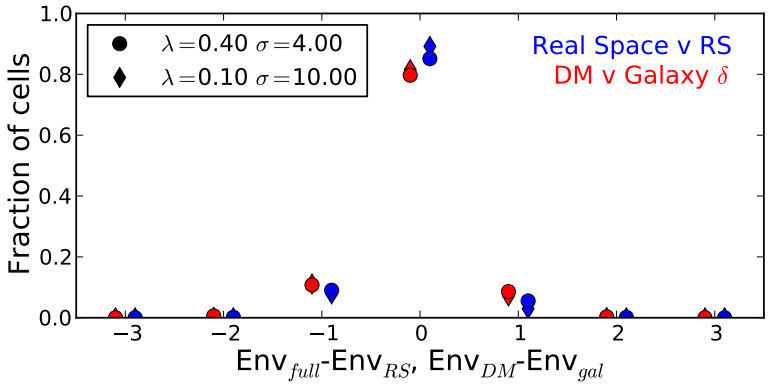


Figure 3.14 The cells that are classified differently, with **Diamonds:** (σ, $\lambda_{\rm th}$)=(10 h^{-1}Mpc, 0.1), **circles:** (σ, $\lambda_{\rm th}$)=(4 h^{-1}Mpc, 0.4), when **red:** the density field is derived from the distribution of galaxies rather than the underlying dark matter density field, or **blue:** when galaxies in redshift- rather than real-space are used. The x-axis represents the difference in the number of eigenvalues above the threshold between the classifications, averaged over three realisations of a GAMA field (with a small offset from the discrete values for clarity).

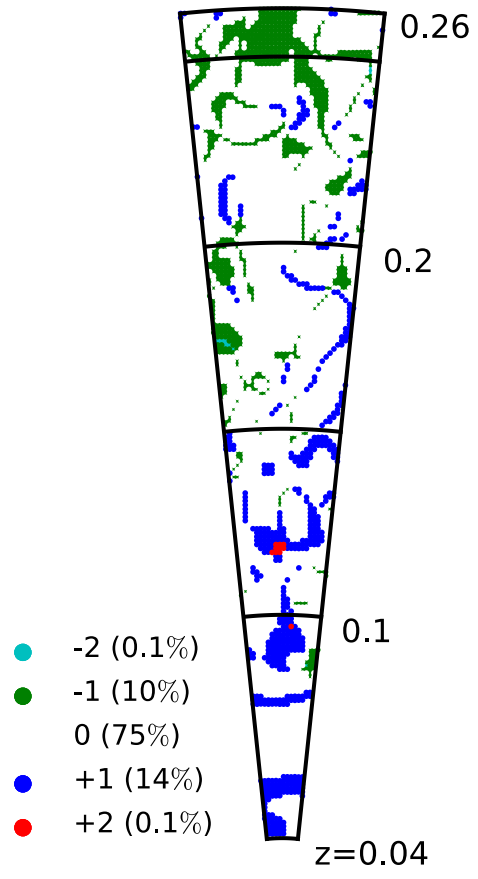


Figure 3.15 A test of the effects of limited information in observational catalogues on resulting geometric environment classifications. Full-information results are computed from the underlying DM density field using the full periodic $1h^{-1}$Gpc simulation. Limited-information results use galaxies from the simulation, in redshift space, discarding information outside of a volume representative of a GAMA field and implementing the reflection technique described in the text. Coloured regions of the figure show the cells that are classified differently between the two approaches for an example realisation of a GAMA field and (σ, $\lambda_{\rm th}$)=(10 h^{-1}Mpc, 0.1). Colour code in the key refers to the difference, ΔN, in the number of eigenvalues above the threshold, N^{+}, between the full simulation and the limited-information survey classifications, e.g. $\Delta N = N^{+}_{\rm FULL} - N^{+}_{\rm LIM}$. Hence each cell has a discrete value of ΔN, with $-3 \leq \Delta N \leq 3$. The percentage of all cells with a given ΔN, within three realisations, is indicated in the key.

causes of error discussed above are all introduced.

3.3 Observed cosmic web within GAMA

A density field is constructed from the GAMA galaxies detailed above for each of the three equatorial fields separately. As discussed in Section 3.2.4, the volume immediately outside the survey region is populated with cloned galaxies reflected along the boundaries of each field. This is in order to reduce the effects of the survey geometry when smoothing the density field.

Fig. 3.16 illustrates the main steps in the classification of one of the GAMA fields. The galaxies are interpolated onto a Cartesian grid and an overdensity field generated by comparison with the catalogue of randomly positioned cloned galaxies. This overdensity field is smoothed by a Gaussian window function of width σ_s in Fourier space. Following that, the potential and its second derivatives are calculated, from which the three eigenvalues can be derived for each cell of the grid. The dimensionality of collapse is approximated by the number of eigenvalues above the chosen eigenvalue threshold and from this each cell is classified as either a void, sheet, filament or knot. Finally, the galaxy catalogue is split into four environmentally defined subcatalogues by assigning the galaxies the geometric environment of the cell in which they reside.

Fig. 3.16a plots the distribution of all galaxies within $\pm 1^\circ$ of the central declination of the G12 field. Fig. 3.16b is the resulting density field along the central declination of the G12 field, after smoothing with an effective smoothing scale of $\sigma = 10\,h^{-1}$Mpc. Fig. 3.16c and Fig. 3.16d show the resulting geometric environments of the central declination slice of the G12 field for the two parameter sets, $(\sigma, \lambda_{\rm th}) = (10\,h^{-1}{\rm Mpc}, 0.1)$ and $(4\,h^{-1}{\rm Mpc}, 0.4)$ respectively. As may be expected, both figures display a similar basic skeleton, with the larger smoothing scale resulting in larger geometric structures. The knots in particular are visibly larger in Fig. 3.16c.

3.3.1 Geometric environments of GAMA galaxies and groups

The geometric environments of all galaxies and groups in the sample are defined by the environment of the cell to which they belong. Fig. 3.17 and Fig. 3.18

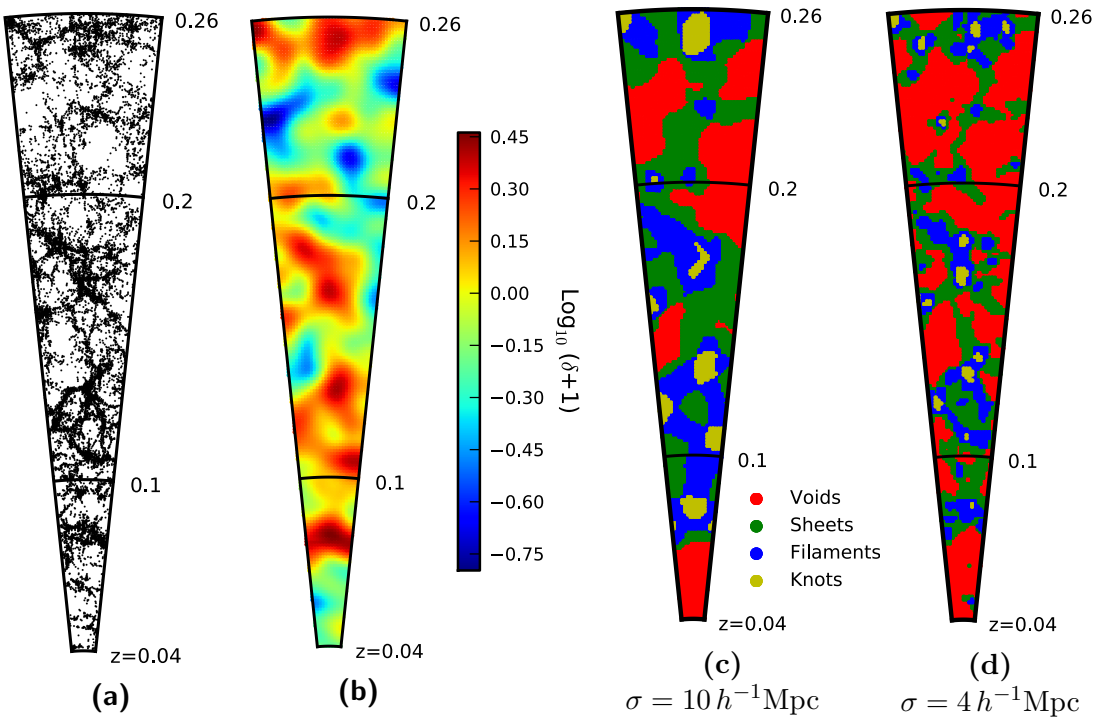


Figure 3.16 An example of the classification of geometric environments within the GAMA G12 field. **(a)** Distribution of galaxies within $\pm 1^\circ$ of the central declination. **(b)** The density contrast field, δ, derived from **(a)** after interpolation of the galaxies on to a Cartesian mesh and smoothing with a Gaussian kernel of effective width $\sigma = 10\,h^{-1}\mathrm{Mpc}$, with a colour scale proportional to $\log_{10}(\delta + 1)$ as given by the colour bar to the right. **(c)** The resulting geometric environment classifications, with an eigenvalue threshold of $\lambda_{\mathrm{th}} = 0.1$, from the smoothed density contrast field in **(b)**. **(d)** The geometric environments for the second parameter set, $(\sigma, \lambda_{\mathrm{th}}) = (4\,h^{-1}\mathrm{Mpc}, 0.4)$. Environments are colour coded as shown in the key, e.g.: red, green, blue and yellow for voids, sheets, filaments and knots respectively. Whilst panel **(a)** shows a 2D projection of galaxies, the slices shown in panels **(b)**, **(c)** and **(d)** show the 2D plane of the central declination; they show the value (density contrast or environment) of whichever cell is intersected by the central declination.

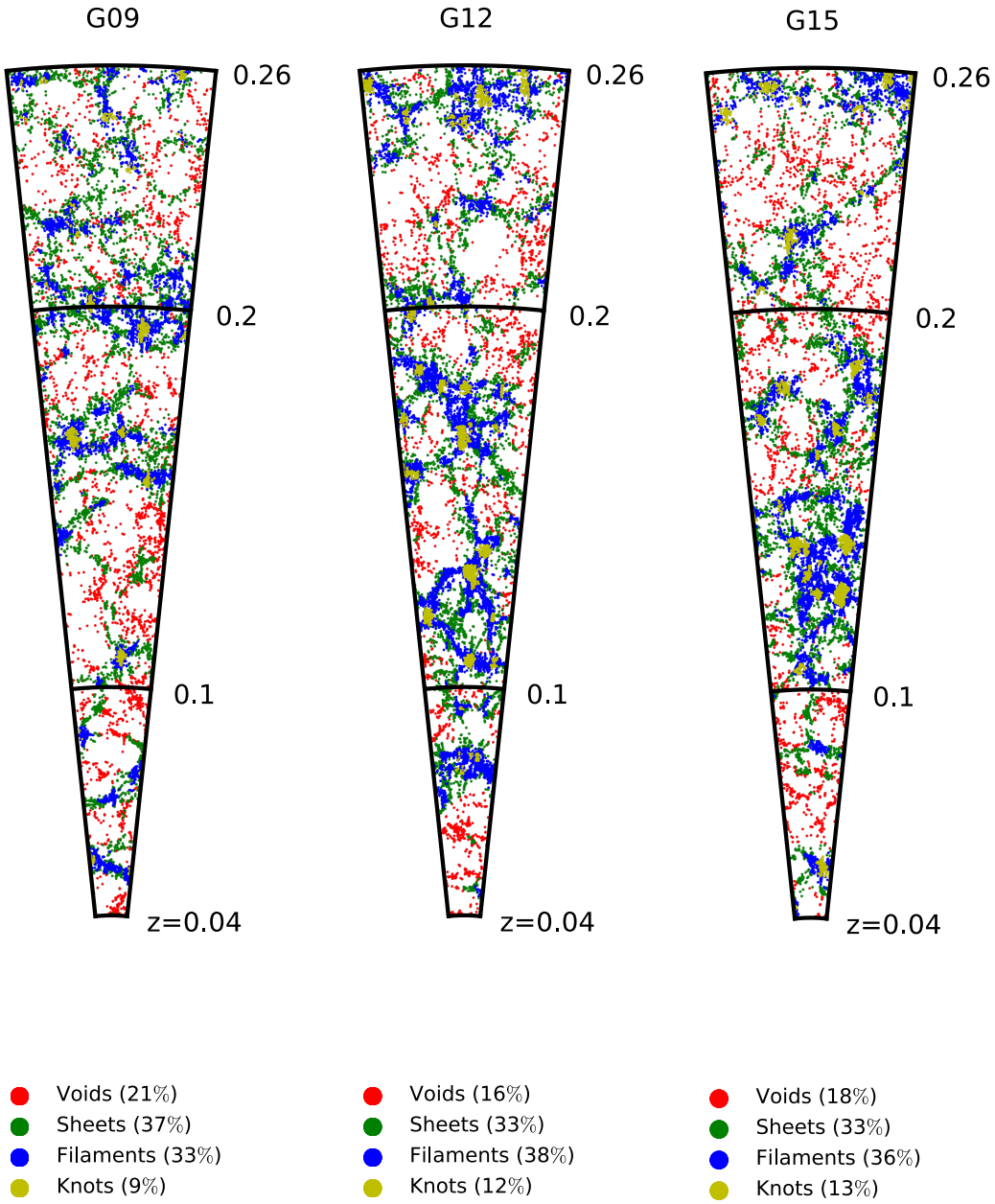


Figure 3.17 The distribution of galaxies in the three equatorial GAMA fields within $\pm 1^{\circ}$ of the central declination. Galaxies are colour coded by their resulting geometric environment classification after smoothing with a Gaussian of width $\sigma = 4\,h^{-1}\mathrm{Mpc}$ and applying a threshold of $\lambda_{\mathrm{th}} = 0.4$. For each of the GAMA fields, the percentage of galaxies within each of the four environments is shown in the keys beneath the cones.

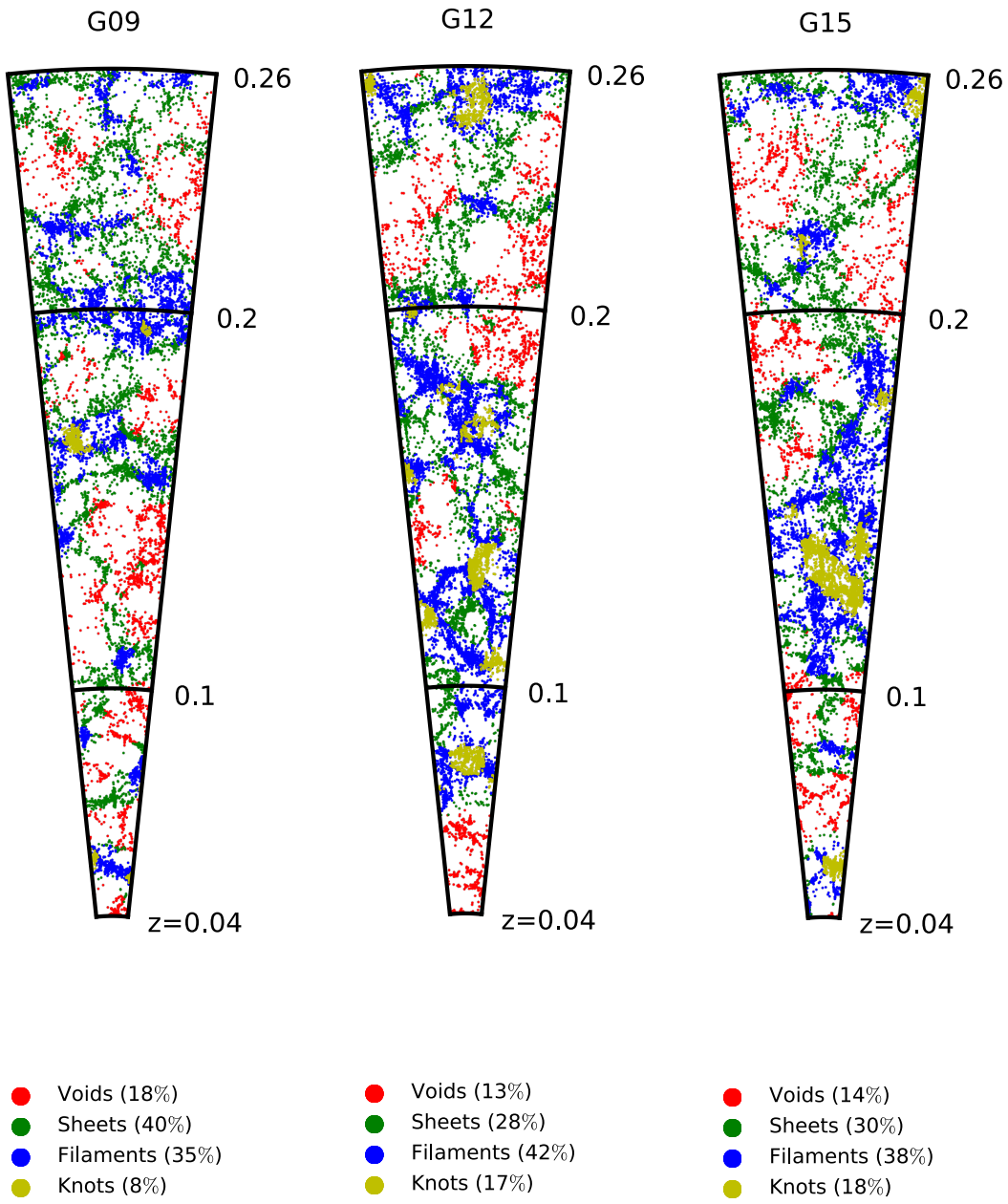


Figure 3.18 The distribution of galaxies in the three equatorial GAMA fields within $\pm 1^{\circ}$ of the central declination. Galaxies are colour coded by their resulting geometric environment classification after smoothing with a Gaussian of width $\sigma = 10\,h^{-1}\mathrm{Mpc}$ and applying a threshold of $\lambda_{\mathrm{th}} = 0.1$. For each of the GAMA fields, the percentage of galaxies within each of the four environments is shown in the keys beneath the cones.

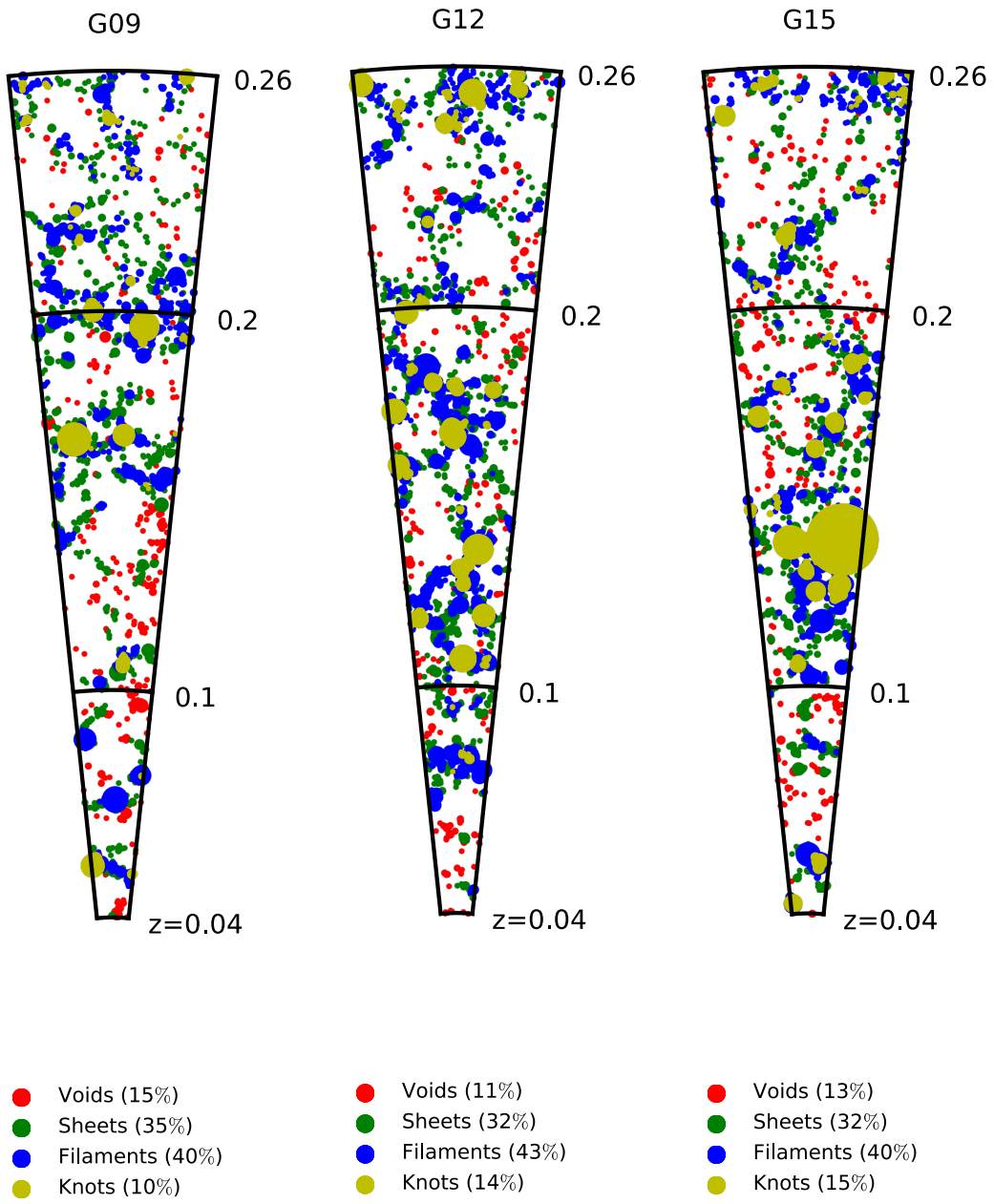


Figure 3.19 The distribution of groups in the three equatorial GAMA fields within $\pm 1^{\circ}$ of the central declination. Groups are colour coded by their resulting geometric environment classification after smoothing with a Gaussian of width $\sigma = 4\,h^{-1}\mathrm{Mpc}$ and applying a threshold of $\lambda_{\mathrm{th}} = 0.4$, the size of the point for each group is proportional to the number of members in the group. For each of the GAMA fields, the percentage of groups within each of the four environments is shown in the keys beneath the cones.

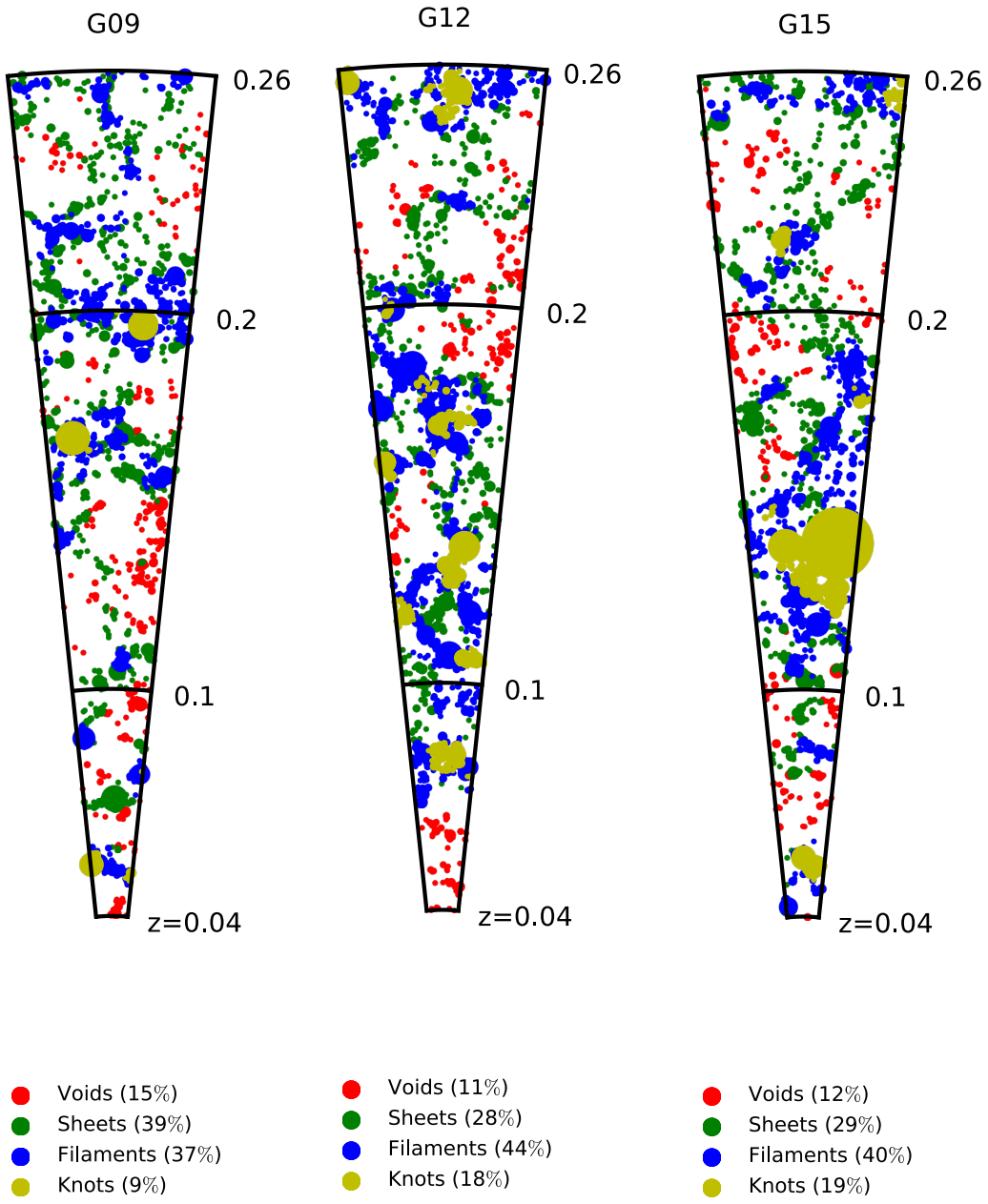


Figure 3.20 The distribution of groups in the three equatorial GAMA fields within $\pm 1^{\circ}$ of the central declination. Groups are colour coded by their resulting geometric environment classification after smoothing with a Gaussian of width $\sigma = 10\,h^{-1}\mathrm{Mpc}$ and applying a threshold of $\lambda_{\mathrm{th}} = 0.1$, the size of the point for each group is proportional to the number of members in the group. For each of the GAMA fields, the percentage of groups within each of the four environments is shown in the keys beneath the cones.

show the geometric environment classifications of galaxies around the central declination slice of each of the three GAMA fields for environments defined with the parameter sets $(\sigma, \lambda_{\rm th}) = (4\,h^{-1}{\rm Mpc}, 0.4)$ and $(10\,h^{-1}{\rm Mpc}, 0.1)$ respectively. Though the sheets are not visually well captured in a 2D figure, the galaxies in filaments stand out clearly, particularly when the filament happens to lie in the plane of the figure[1]. The void galaxies are in less populated regions but sometimes exhibit small-scale clustering which has been smoothed out during the filtering process. Similarly, Fig. 3.19 and Fig. 3.20 show the geometric environment classifications of groups around the central declination slice of each of the three GAMA fields for the two parameter sets. The size of the pointer for each group is proportional to the number of members of the group. It can be seen that the larger groups are predominantly found in knots, and the groups in void regions tend to have fewer members. The existence of this trend is reinforced by the cumulative histograms shown in Fig. 3.21. The average number of members of a group increases between void and knot environments, as may be expected.

As well as having a distinct shape, the different geometric environments are also strongly distinct in terms of density. The distributions of local overdensities of the galaxies within each geometric environment are shown in Fig. 3.22, where the overdensity is calculated from the number of galaxies within a sphere of radius $8\,h^{-1}$Mpc centred on the location of each galaxy rather than the grid-based overdensity measure used during the environment classification process. This density measure follows from the work of Croton et al. (2005) and is chosen for consistency with McNaught-Roberts et al. (2014); it involves selecting a 'density-defining-population' of galaxies, which constitute a volume-limited sample over the redshift range of interest (as depicted by the blue box in Fig. 3.6). All galaxies within the $8\,h^{-1}$Mpc radius contribute to the density measure if they are part of the density-defining-population, including the galaxy for which the density is being measured. The measured densities are converted to δ_8, a measure of overdensity, by comparison with the effective mean density within the sample.

As expected, the average overdensity increases as the dimensionality of the environment decreases (note that the 3D voids are the highest dimension of environment, with knots considered to have the lowest dimensionality, having collapsed in all dimensions); most void galaxies are found in underdense regions,

[1]For an animated view of the 3D distribution of geometric environments, see `http://www.roe.ac.uk/~ee/GAMA`

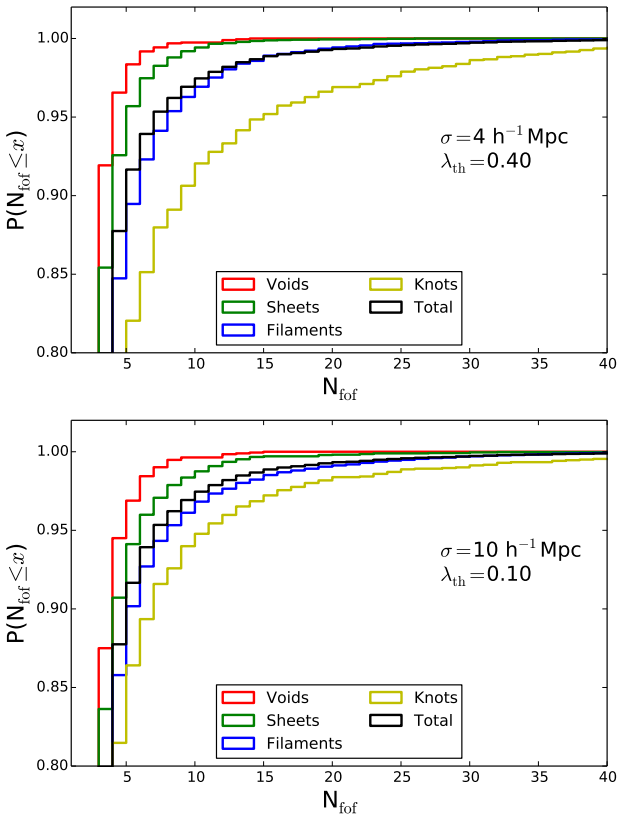


Figure 3.21 Cumulative distribution of number of group members, N_{fof}, for groups split by geometric environments derived from the two parameter sets: $(\sigma, \lambda_{th}) = (4\,h^{-1}\mathrm{Mpc}, 0.4)$ and $(10\,h^{-1}\mathrm{Mpc}, 0.1)$.

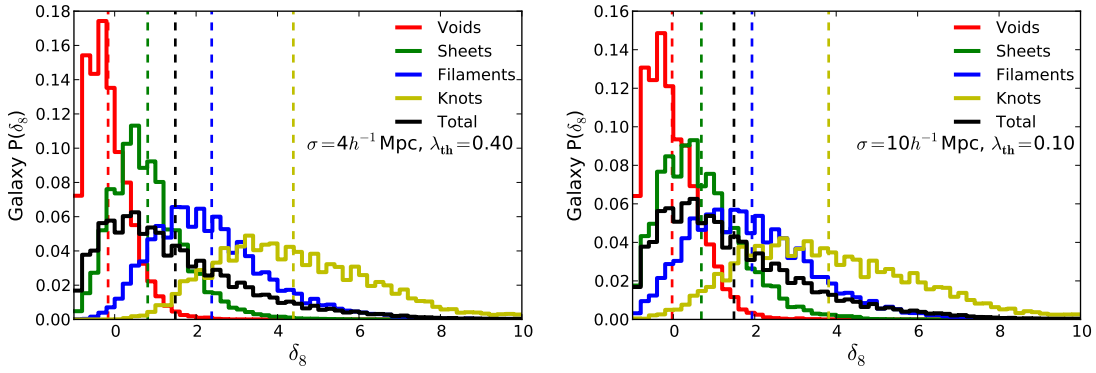


Figure 3.22 Distribution of local overdensities for galaxies split by geometric environment. Dashed lines indicate the average overdensities, as given in Table 3.2, of all galaxies within each environment. The overdensity, δ_8, is derived from the number of galaxies within a sphere of radius $8\,h^{-1}\mathrm{Mpc}$. The overdensity increases as the dimension of the environment reduces; but because there is a wide range of overdensities in each environment it is possible to look for a dependence of galaxy properties on δ_8 and geometric environment separately.

almost no knot galaxies reside in underdense regions and instead live in highly overdense areas. One may be surprised that there is a significant proportion of voids which are found to be slightly overdense. This is not the first time this definition of geometric environments has produced such a result (some voids identified in the analysis of simulated data in Chapter 2 were found in overdense areas). If the classification was forced to produce only underdense voids, whilst retaining the simplicity of a classification based on the three eigenvalues of the tidal tensor only, it would be necessary to reduce the eigenvalue threshold until this condition was satisfied. Choosing such a low value for the threshold contradicts with the intentions described in section 3.2.3, where the need for good statistics and a near-comparable number of objects in each environment in order to search for small modulations was discussed. Such a low threshold also fails to produce a good visual impression of the web, resulting in apparently 3D regions being classified as a 2D sheet. The broad distribution of densities within each environment shows the geometric environment holds more information than the density alone. Environments derived from the $10\,h^{-1}$Mpc smoothed density field show a slightly larger spread of densities in any given environment than for the $4\,h^{-1}$Mpc field, though the distributions are relatively similar.

3.3.2 Other GAMA LSS analyses

An alternative method of classifying LSS within the GAMA regions was conducted by Alpaslan et al. 2014 (hereafter A14). When the results of A14 are compared to those presented here, a strong visual agreement for the filamentary regions is apparent but the agreement in the classification of voids is more limited. With the somewhat flexible definitions of geometric environments this is neither a surprise nor cause for concern. On the contrary, it illustrates the variety of meanings of terms such as voids, even within the context of the cosmic web. Hence, when interpreting any results in the context of the web it is important for one to have a clear quantitative understanding of the how the environments in question are defined.

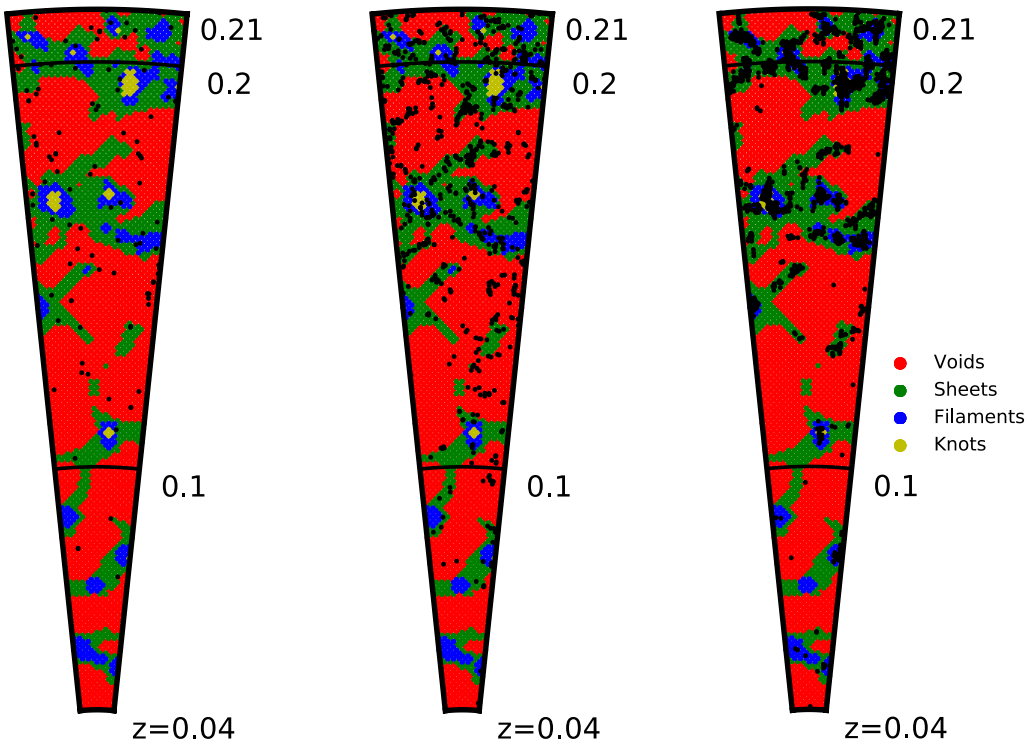
A14 implemented a minimal spanning tree algorithm to identify 643 filaments within the same three GAMA equatorial regions used in this work, with a slightly lower redshift cut of $z < 0.213$. A14 also identified a secondary population of smaller coherent structures ('tendrils') and a population of isolated void galaxies.

Fig. 3.23 shows the central declination slice of the G9 field, the geometric environments as classified by this work, and all objects in each of the three populations as identified in A14 within $\pm 0.5^\circ$ of the central declination. The filaments of A14 are found to be visually consistent with the filamentary regions identified in this work. The tendrils and voids of A14 favour the underdense environments of voids and sheets. Note that the results for environments computed with $\sigma = 4\,h^{-1}$Mpc are shown, a similar scale to the $r = 4.13\,h^{-1}$Mpc used in A14 as the maximum distance allowed between a galaxy and a filament. One might suggest that the 'void galaxies' as identified in A14 should be thought of as isolated galaxies, whereas the voids as defined in this work correspond to larger geometric structures. A more quantitative comparison is presented in Fig. 3.24; the histograms illustrate, for each A14 population, the number of galaxies belonging to each of the geometric environments identified in this work through the application of the tidal tensor prescription. The dashed lines in the figure indicate the number of galaxies in the GAMA sample in each environment as defined by this work, normalised by the size of the each A14 population, hence the dashed lines indicate the proportion which would be expected from a purely random selection.

3.4 Summary

This chapter has presented a detailed analysis of the application of the tidal tensor prescription to data from the GAMA spectroscopic redshift survey. The key achievements of this chapter can be summarised as follows:

- The redshift dependence of the tidal tensor prescription method for classifying geometric environments was investigated. The variation of the bias and growth factor, which act to modify the effective eigenvalue threshold, was estimated, leading to a model for a redshift dependent threshold. This variable threshold was tested on the GAMA dataset and the fraction of the survey classified as each of the four geometric environments within redshift bins out to $z = 0.5$ was measured and compared to the results of a constant threshold. These investigations motivated the choice to focus on only the lower redshift regime of GAMA, namely $0.04 < z < 0.26$. A volume limited sample of galaxies was extracted and



(a) A14 Voids **(b)** A14 Tendrils **(c)** A14 Filaments

Figure 3.23 A comparison of large scale structure identified by this work, and by A14, within the central declination of the G9 field. The geometric environments defined by the tidal tensor prescription, calculated with $\lambda_{\rm th} = 0.4$ and $\sigma = 4\,h^{-1}{\rm Mpc}$, are shown by the background colours with red, green, blue and yellow indicating voids, sheets, filaments and knots respectively. From left to right the black dots in the figures show the positions of all galaxies within $\pm 0.5^{\circ}$ of the central declination in the A14 populations voids, tendrils and filaments respectively.

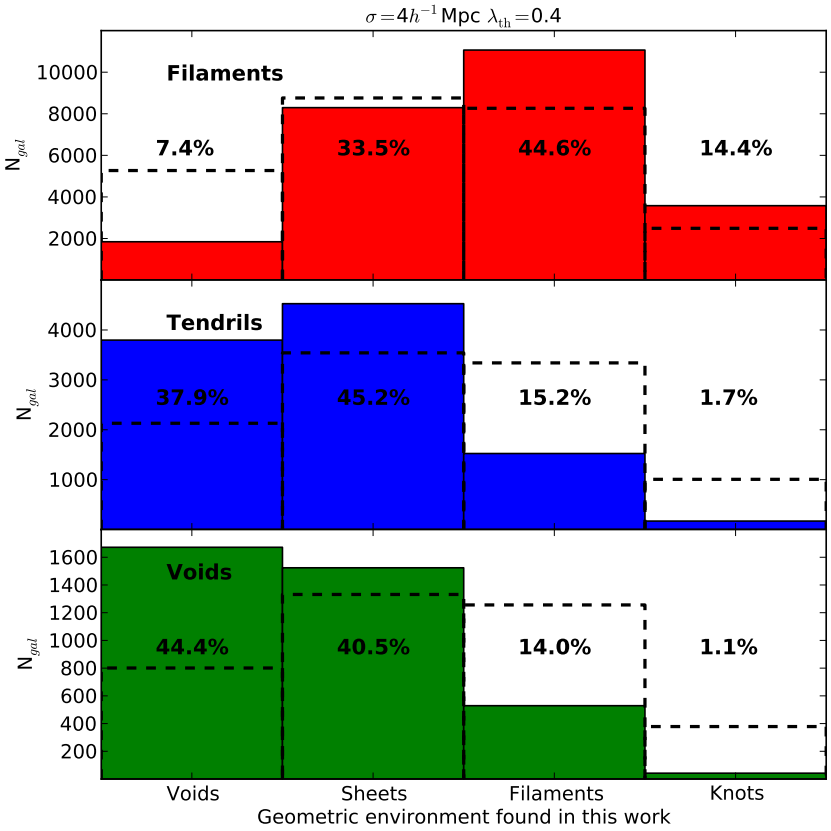


Figure 3.24 Comparison between large scale structure identified in A14 and in this work. Each row shows how either the filaments, tendrils or voids identified in A14 are classified in this work, with $\lambda_{\rm th} = 0.4$ and $\sigma = 4\,h^{-1}{\rm Mpc}$. The percentages given in the figure show, for each A14 population, the percentage of galaxies in each of the environments in this work. Dashed lines indicate the number of galaxies in the full GAMA sample classified by the tidal tensor prescription in each geometric environment, normalised by the number of galaxies in the A14 population which each row represents. Hence the dashed lines, which are the same for each panel before normalisation, can be thought of as the expected distribution of a random selection from all galaxies.

the resulting geometric environments compared to the magnitude limited sample, showing no strong benefit to restricting the sample to cover a constant magnitude range.

- The MDR1 simulated dataset was used to quantify the inaccuracies introduced when using a limited and imperfect dataset such as GAMA. A reflection technique was introduced and tested with the simulated data and was shown to reduce the effects of the survey geometry. The effect of using galaxies in redshift space rather than the underlying dark matter in order to estimate the gravitational potential was also investigated using the simulated dataset. It was found that these limitations change the geometric environment classifications of $< 25\%$ of the volume.

- Two optimal parameter sets were chosen by requiring a comparable split of galaxies between the environments. The geometric environments of the GAMA volume, galaxies and groups were presented for each of these two parameter sets for the three equatorial fields of the GAMA survey. These environments were compared to an alternative method of LSS identification within GAMA.

With an appropriate sample identified within the GAMA survey, and the geometric environments classified by application of the tidal tensor prescription, in the following two chapters the direct influence of the tidal field on large-scale structure is investigated.

Chapter 4

Large-scale structure within the cosmic web

With the advent of numerical simulations, and of expansive and detailed galaxy surveys, we are able to test in more detail the extent to which different properties of the environment may influence LSS formation. For example, Hahn et al. (2009) find the mass assembly history of halos to be influenced by tidal effects, and note that tidal suppression of small halos may be especially effective in filamentary regions. Ludlow & Porciani (2011) used cosmological ΛCDM simulations to test the central ansatz of the peaks formalism, in which halos evolve from peaks in the linear density field when smoothed with a filter related to the halos characteristic mass. Although they found the majority of halos to be consistent with this picture, they identify a small but significant population of halos showing disparity and find these halos are, on average, more strongly compressed by tidal forces.

The visible manifestation of such tidal forces is the striking way in which gravitational instability rearranges the nearly homogeneous initial density field into the cosmic web. Recent years have seen an increased interest in methods of classifying the cosmic web (see e.g. Cautun et al. 2013 for an overview). Many of these studies have been applied to numerical simulations, finding some promising detection of LSS alignments with filaments (Codis et al. 2012, Forero-Romero et al. 2014). Studies of geometric environments in observational datasets have more often focused on identifying individual structures such as voids or filaments rather than classifying the global volume.

The previous chapter presented an application of the tidal tensor prescription,

based on the second derivatives of the gravitational pseudo-potential, to the Galaxy And Mass Assembly (GAMA) spectroscopic redshift survey. The surveyed volume was classified as either a void, a sheet, a filament or a knot by approximating the dimensionality of collapse. Building on this, it is possible to investigate various properties of LSS as a function of location within the cosmic web. The motivation is to search for any correlation of LSS properties with this non-local aspect of the density field and search for any modulation that the geometric environment introduces. Of course, some properties may be affected in a completely local manner (see e.g. Wijesinghe et al. 2012, Brough et al. 2013 and Robotham et al. 2013 for previous studies of the dependence of GAMA galaxy properties on local environments), so in parallel it is important to track the dependences that are purely functions of overdensity.

This chapter is structured as follows: In Section 4.1 the galaxy luminosity function within the GAMA survey is studied. The conditional luminosity function is measured within geometric environments and the direct influence of the web is tested by comparing the measurement with luminosity functions measured for galaxies with matching local density distributions. In Section 4.2 a similar analysis is applied to the GAMA group mass function. The global group mass function within the GAMA sample is investigated along with its dependence on location within the cosmic web. This allows the main conclusion of Chapter 2, that the halo mass function is independent of geometric environment for a given local overdensity, to be investigated within the geometric environments identified within GAMA.

4.1 Luminosity functions and geometric environment

The galaxy luminosity function (LF) is central to studies of galaxy formation and evolution. A strong dependence on local environment of many galactic properties, such as morphology, star formation rate and colour, has long been established (e.g. Dressler 1980, Gómez et al. 2003, Balogh et al. 2004). However, many models of galaxy formation assume only a very limited environmental impact. In standard halo-occupation models and some semi-empirical models, galaxy properties are assumed to depend only upon the mass of the host halo

or its merger history (Kauffmann et al. 1993, van den Bosch et al. 2007). With the existence of ever larger spectroscopic redshift surveys, such as SDSS and 2dFGRS, we are able to test these basic assumptions and search for evidence suggesting more complicated models. For example, a dependence of the galaxy LF on local density has been investigated and the LF has been shown to vary smoothly with overdensity, brightening continuously from void to cluster regions with no significant variation in the LF slope (Croton et al. 2005, McNaught-Roberts et al. 2014). Guo et al. (2014) measured the satellite LF of primary galaxies in SDSS and found a significant difference between galaxies residing in filaments and those that do not, suggesting that the filamentary environment has a direct effect on the efficiency of galaxy formation.

There are many physical mechanisms that may be involved in determining the galaxy LF: mergers, tidal interactions and ram pressure gas stripping for example may all affect the luminosity of galaxies and induce an environmental dependence. Certainly some of these mechanisms must be influenced by the local matter density, purely through its impact on the population of dark-matter halos – which in turn affects the properties of the galaxies hosted by the halos (Vale & Ostriker 2004, Moster et al. 2010). Much theoretical work concerning the formation, clustering and mass distribution of dark matter halos has already been undertaken. For example, the standard explanation for biased galaxy clustering uses the peak-background split formalism (Bardeen et al. 1986, Cole & Kaiser 1989), in which the large-scale density field modulates the likelihood of collapse of halos. But beyond this, it is conceivable that some galaxy properties may be linked not only with overdensity, but with geometric environment, and it is this possibility which is studied in the following subsection.

4.1.1 The GAMA luminosity function

In full, GAMA observes galaxies out to $z \simeq 0.5$ and $r < 19.8$ mag; but in this work, as discussed in Section 3.1.1, only the lower redshift regime is studied. For consistency with the previous analysis of the environmental dependence of the luminosity function within GAMA by McNaught-Roberts et al. (2014) (hereafter MNR14), a sample of 113000 galaxies satisfying $0.04 < z < 0.263$ selected from the three equatorial regions of GAMA: G09, G12 and G15, each spanning $12^\circ \times 5^\circ$ is used. When testing the effects of the chosen sample, no benefit was found to

restricting the catalogue to a volume-limited subset, hence no absolute magnitude cuts are imposed. All galaxies with a GAMA redshift quality rating of `nQ` > 2, indicating the redshift is sufficiently reliable to be included in scientific analyses, and an appropriate visual classification flag (`VIS_CLASS` $= 0, 1$ or 255, Baldry et al. 2010) are used.

The galaxy LF is measured independently for each geometric environment as presented in Section 3.3, using k-corrected and luminosity evolution corrected absolute r-band magnitudes, M_e^r, and following the approach taken in MNR14. This method adopts the step-wise maximum likelihood estimator (Efstathiou et al. 1988), and normalises the LF taking into account the effective fraction of the volume covered by a given geometric environment ($f_{\rm env}$):

$$N = \Omega \int_{z_1}^{z_2} \mathrm{d}z \frac{\mathrm{d}V}{\mathrm{d}z\mathrm{d}\Omega} \int_{M_{\rm faint}(z)}^{M_{\rm bright}(z)} \frac{1}{f_{\rm env}} \phi(M)\, \mathrm{d}M \tag{4.1}$$

for a galaxy sample with number of galaxies (N), redshift limits (z_1 and z_2) and solid angle (Ω). The resultant conditional luminosity functions for each geometric environment are shown in Fig. 4.1 with jackknife error bars. These conditional LFs reveal a higher number density of luminous galaxies in lower dimensional environments, introducing a vertical shift of the LF. A Schechter function (Schechter 1976), given by

$$\phi(M) = \frac{\ln 10}{2.5} \phi^* 10^{0.4(M^*-M)(1+\alpha)} \exp\left(-10^{0.4(M^*-M)}\right), \tag{4.2}$$

is used to characterise the magnitude and shape of the luminosity function. Here α describes the power law slope of the faint end, M^* describes the magnitude at which there is a break from the power law, or the 'knee' of the LF, and ϕ^* describes the normalisation. The solid lines of Fig. 4.1 show best fitting Schechter functions for each LF, with the best fitting parameters given in Table 4.1. There is a clear increase in the normalisation of the LF from voids to knots, shown by the steady increase of ϕ^*. The turnover point, M^*, of the LF moves towards brighter magnitudes from voids to knots, suggesting brighter galaxies have an increased bias towards lower dimensional regions. Note that there may be some environmentally dependent degeneracies in the α and M^* parameters (see Fig. D1 of MNR14).

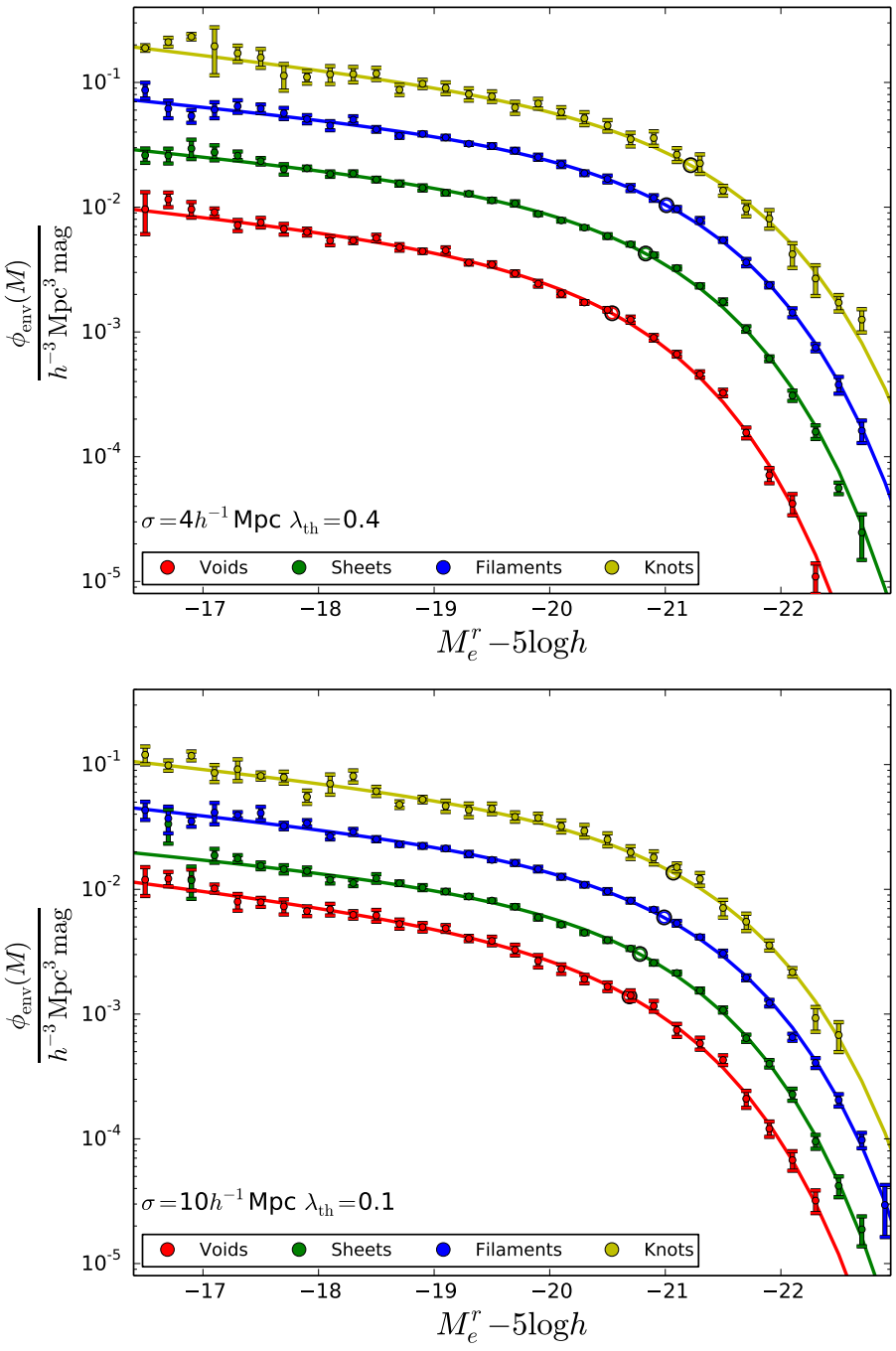


Figure 4.1 The galaxy luminosity functions and corresponding jackknife errors of the 4 subcatalogues produced by splitting the GAMA sample according to geometric environment, defined with $\sigma = 4$ or $10\,h^{-1}{\rm Mpc}$. Solid lines show best fitting Schechter functions for each conditional LF, open circles show the best fitting value of M^*, given in Table 4.1. The normalisation, or ϕ^* in a Schechter function fit, is seen to increase significantly between high- and low-dimensional environments.

Environment	$\sigma(\,h^{-1}\mathrm{Mpc})$	α	$\log_{10}[\phi^*/\,h^{-3}\mathrm{Mpc}^{-3}]$	$M^* - 5\log h$
Voids	4	-1.25 ± 0.02	-2.38 ± 0.02	-20.54 ± 0.03
Sheets	4	-1.23 ± 0.02	-1.90 ± 0.04	-20.83 ± 0.04
Filaments	4	-1.22 ± 0.02	-1.51 ± 0.04	-21.01 ± 0.05
Knots	4	-1.27 ± 0.06	-1.19 ± 0.14	-21.22 ± 0.14
Voids	10	-1.29 ± 0.03	-2.39 ± 0.06	-20.69 ± 0.06
Sheets	10	-1.22 ± 0.02	-2.05 ± 0.03	-20.78 ± 0.03
Filaments	10	-1.24 ± 0.02	-1.75 ± 0.03	-20.99 ± 0.04
Knots	10	-1.25 ± 0.04	-1.40 ± 0.09	-21.07 ± 0.09

Table 4.1 Best-fitting parameters found for a non-linear least squares Schechter function (Eq. 4.2) fit to the conditional LF of each environment, classified with either $(\sigma, \lambda_{\mathrm{th}}) = (4\,h^{-1}\mathrm{Mpc}, 0.4)$ or $(10\,h^{-1}\mathrm{Mpc}, 0.1)$. α shows no clear trend with environment, ϕ^* shows a significant, steady increase from voids to knots and M^* brightens from voids to knots. Errors are calculated from the standard deviation of the resultant parameters for 9 jackknife realisations. Note that there may be some degeneracy between α and M^*.

A comparison of the upper and lower panels in Fig. 4.1 shows the impact of the choice of different smoothing scales and thresholds. Using a range of parameters following the optimal black curve in Fig. 3.9, it was found that the magnitude of the difference between the conditional LFs increases as the smoothing scale decreases or as the eigenvalue threshold decreases. This tends to introduce only a vertical shift to the functions, characterised by ϕ^*, whilst the shapes of the luminosity functions do not show significant dependence on the smoothing and threshold parameters.

In order to remove some of the vertical offset and illustrate the difference in shape between the LFs of each geometric environment, Fig. 4.2 plots the ratio of the conditional LFs to a set of scaled reference Schechter functions. The reference function, $\phi_{\mathrm{ref,tot}}$, is found by fitting a Schechter function to all galaxies in the sample. A normalisation is applied to each environment to produce the scaled reference Schechter functions, $\phi_{\mathrm{ref,env}}$, given by

$$\phi_{\mathrm{ref,env}} = \frac{(1 + \bar{\delta}_8^{\mathrm{env}})}{(1 + \bar{\delta}_8^{\mathrm{tot}})} \times \phi_{\mathrm{ref,tot}}, \tag{4.3}$$

where $\bar{\delta}_8^{\mathrm{env}}$ is the average overdensity within an $8\,h^{-1}$Mpc sphere centred on each galaxy of the environment and $\bar{\delta}_8^{\mathrm{tot}} = 8 \times 10^{-3}$ is that of all the galaxies in the full sample. The solid lines in Fig. 4.2 show the ratio of the best fitting Schechter

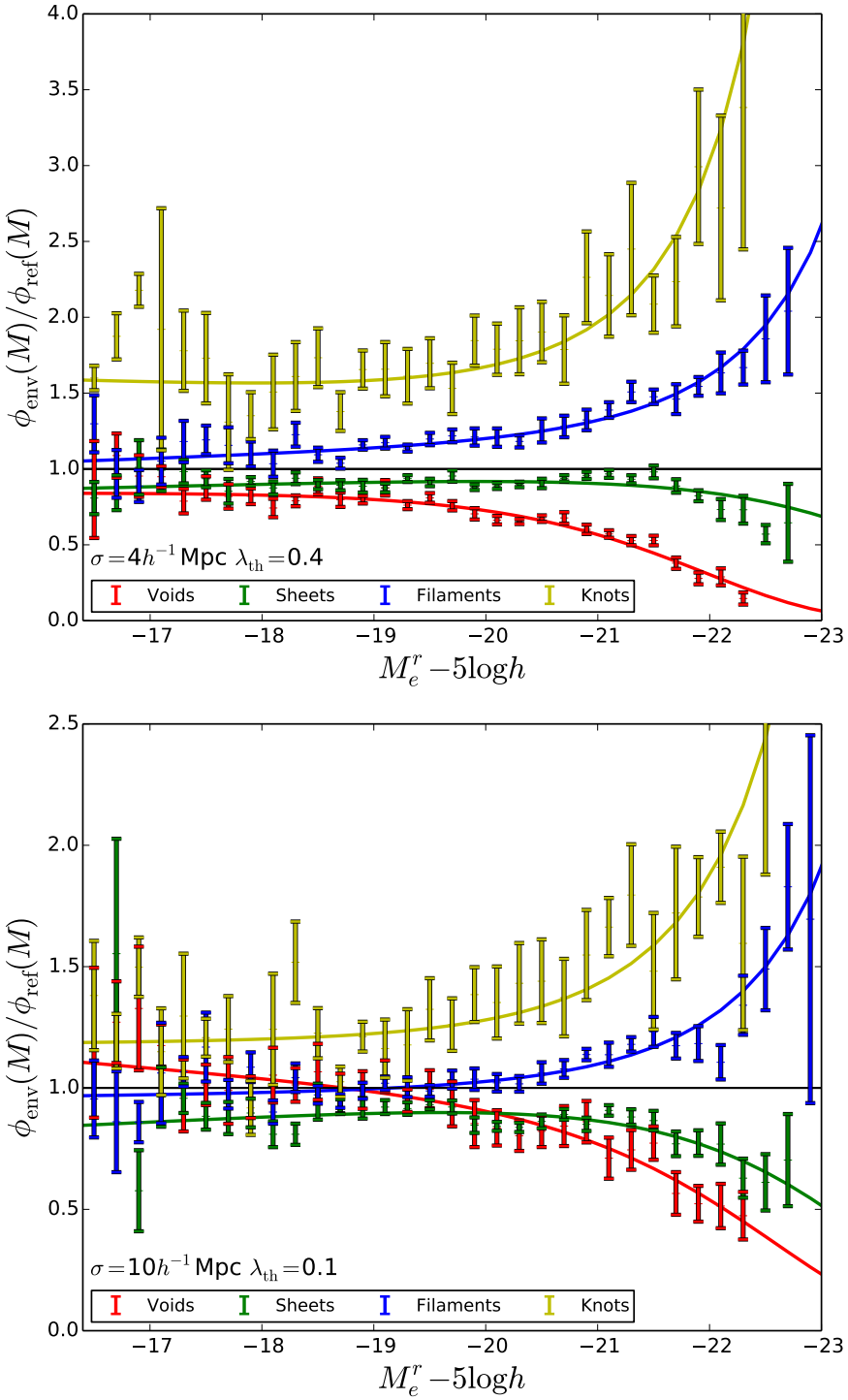


Figure 4.2 Observed environmental luminosity functions (points) and their best fitting Schechter functions (solid) divided by the scaled reference Schechter functions of Eq. 4.3 for each geometric environment, colour coded as shown in the legend. The difference in the shape of the LF between the environments is most apparent at the bright end of the LF, owing to the decrease of the turnover magnitude, M^*, from voids to knots. Note that the linear scaling means that a factor of 2 in excess is more noticeable than a factor of 2 in deficit.

functions for each environment to the reference functions, data points show the ratio for the measured luminosity functions. The departure from the global shape is seen to increase towards the bright end of the LF; the number density of void galaxies decreases as we move towards brighter magnitude bins faster than that of the global population whereas we see a slower decline with brightness for the knot galaxies. The remaining vertical offset is likely to be due to the approximations used in defining $\phi_{\mathrm{ref,env}}$: for example, the effect of bias is not considered, i.e. galaxies are biased tracers of the underlying dark matter density field and the degree of bias may vary between environments.

4.1.2 Direct dependence on geometric environment

It is clear from the results of MNR14 and others (e.g. Hütsi et al. 2002, Croton et al. 2005, Tempel et al. 2011) that local density plays a significant role in determining the number density of luminous galaxies. This work aims to test whether the geometric environment plays any *additional* role. Is it correct to assume that the LF, given a certain local overdensity, will be the same regardless of location within the cosmic web? The analytic results of the dependence of the Schechter fit parameters on local density, $\phi(M|\delta_8)$, presented in MNR14 could be used to answer this question; however, statistical uncertainties in the fit parameters at the extremes of δ_8 limit this approach. Instead the question is addressed by sampling the galaxies in such a way as to remove any additional geometric information from the environment-split subcatalogues, whilst retaining the distribution of local densities. The LFs are recalculated for these resampled catalogues with the hypothesis that any direct modulation by geometric environment will present itself as a disparity between these results.

Four 'shuffled' catalogues are populated by randomly selecting GAMA galaxies from within overdensity bins, such that the δ_8 distribution of each shuffled catalogue matches that of one of the original geometric-environment-split catalogues. This can be thought of as shuffling the galaxies around within regions of the same overdensity (bins of width 0.1 in δ_8 are used). For each galaxy that was included in the original LF, a galaxy with the same local overdensity is picked at random and the original galaxy is effectively replaced with this new one. Thus the overall effect is to remove the geometric environment distinction contained in the original catalogues, whilst maintaining the same distribution of

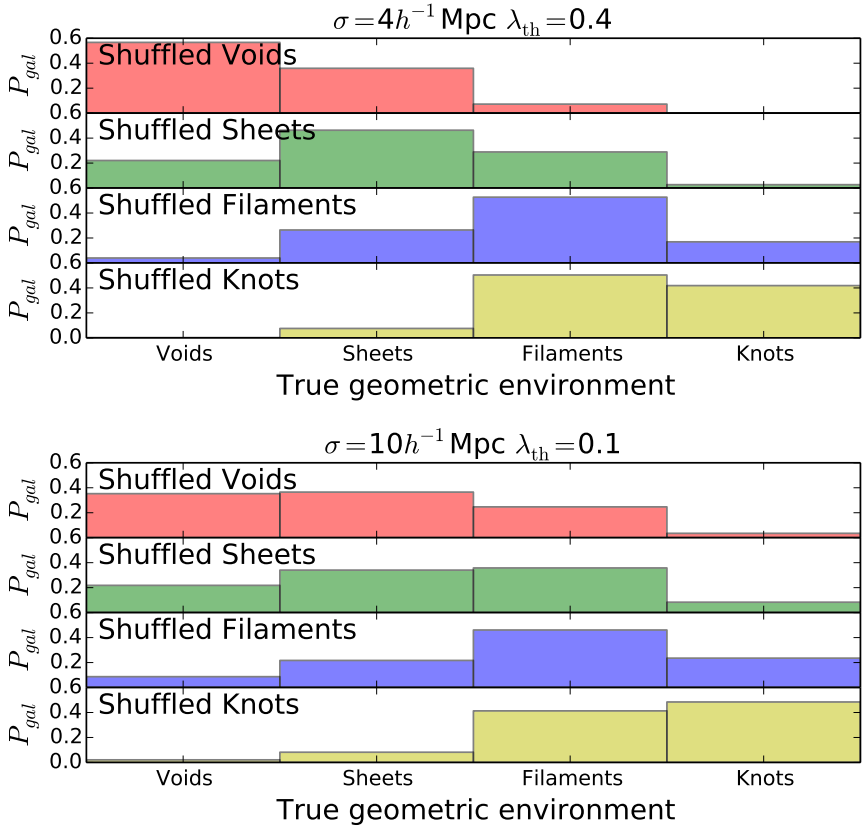


Figure 4.3 The proportion of galaxies from each true geometric environment which were sampled to make up the new shuffled-galaxy catalogues, for $\sigma = 4\,h^{-1}\mathrm{Mpc}$, $\lambda_{\rm th} = 0.4$ (top panel) or $\sigma = 10\,h^{-1}\mathrm{Mpc}$, $\lambda_{\rm th} = 0.1$ (bottom panel). Shuffled catalogues are predominantly composed of galaxies which were also in the original environment catalogue, as is expected from the distribution of overdensities shown in Fig. 3.22. However, there is still significant mixing between environments due to the overlap of the histograms seen in Fig. 3.22, which should change the resultant LF if the geometric environment is having a direct influence.

local densities.

A volume limited sample is required to allow galaxies to be shuffled randomly over different redshifts without moving galaxies out of their observable redshift range. A sample of $\simeq$ 26000 galaxies satisfying $0.021 < z < 0.137$ and $-22 < M_r^e - 5\log h < -18.5$ is used, chosen as a compromise between a large magnitude range and a large sample size. Fig. 4.3 shows the proportion of galaxies in each shuffled catalogue which were taken from each of the four 'true' geometric environments. For example, the bottom row in the top panel of Fig. 4.3 shows that $\simeq 50\%$ of the galaxies in the $\sigma = 4\,h^{-1}$Mpc shuffled-knots catalogue are from a filament environment, and the remaining $\simeq 40\%$ and $\simeq 10\%$ of galaxies were drawn from knot and sheet environments respectively. The combined distribution of densities of this selection of galaxies making up the shuffled-knots catalogue matches the distribution seen in the original knots catalogue. A large fraction of the galaxies in each shuffled catalogue were also in the original geometric environment catalogue, as is expected from the distribution of overdensities, but there is still significant mixing due to the overlap of the histograms seen in Fig. 3.22. Slightly more mixing is seen between the $\sigma = 10\,h^{-1}$Mpc geometric environments due to the slighter broader distribution of densities in any given environment, with more than half the galaxies of each shuffled catalogue being selected from a different geometric environment. The idea is that, if geometric environment has a significant direct effect, we should see different luminosity functions for each shuffled catalogue, in which geometric information is lost, as compared with the initial geometrically-split catalogue. Fig. 4.4 shows the LF for each environment, given by the circles and jackknife error bars, and for an average over 9 realisations of shuffled catalogues, shown by the solid lines. The LFs of the original and the shuffled catalogues are fully consistent, indicating that *the local overdensity is the only significant environmental property affecting the galaxy LF and the cosmic web has no direct influence.* The ratio between the geometric- and the shuffled-LFs, shown in Fig. 4.5, further emphasises that no statistically significant difference between the two measurements is found. The scale dependence of this result was tested by repeating the shuffling process with densities defined over spheres of radii 6 and $12\,h^{-1}$Mpc. No significant differences were found in the results. As a further test, the geometric classifications of individual cells of the initial Cartesian grid were shuffled within density bins

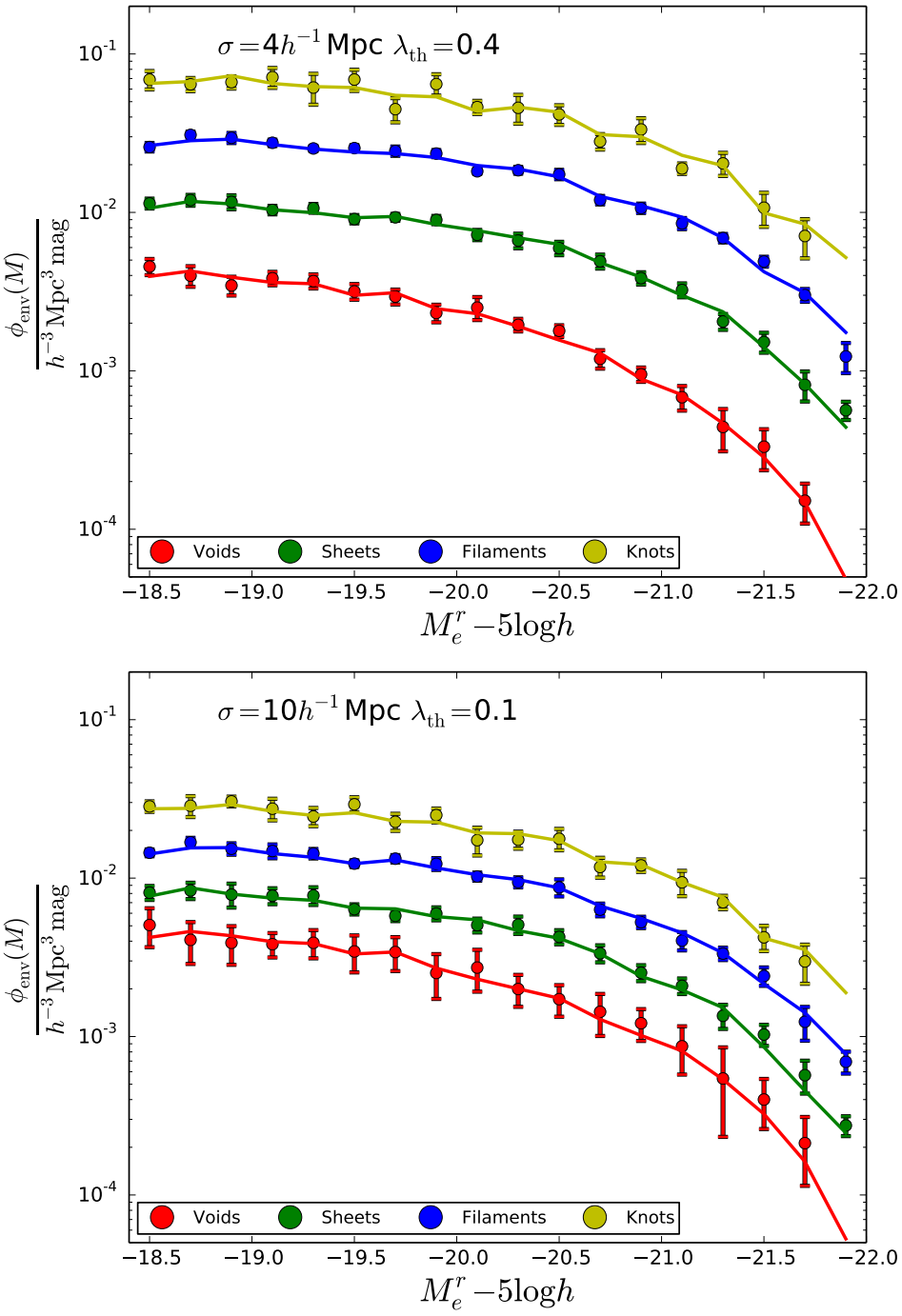
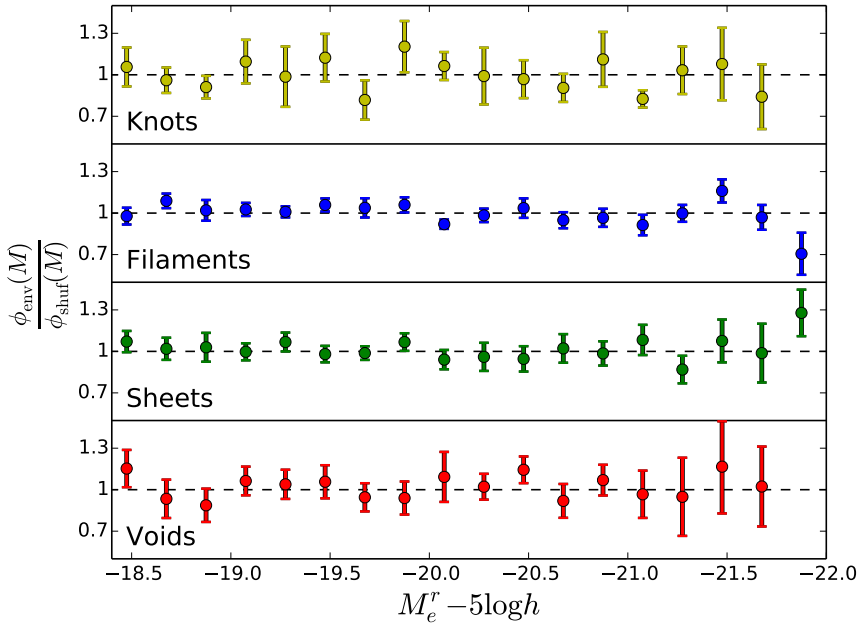
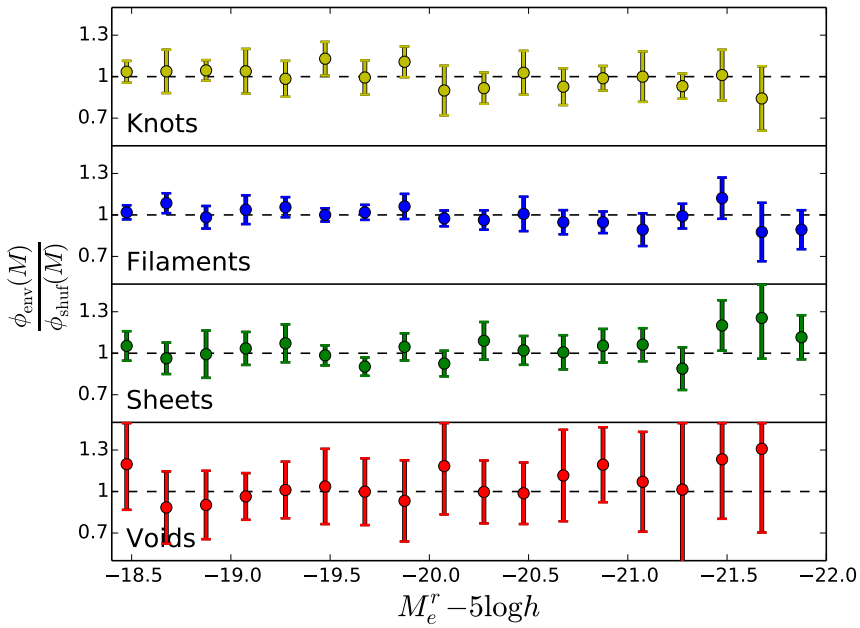


Figure 4.4 The conditional galaxy LFs of the volume limited sample described in section 4.1.2. The LFs for the true geometric environments are indicated by the circle markers, with jackknife error bars. For each geometric environment 9 realisations of shuffled catalogues were created which mimic the distribution of overdensities in the geometric environment but select galaxies randomly regardless of local geometry. The solid lines plot the average of the 9 realisations for each environment and can be seen to be fully consistent with the original LFs, indicating the galaxy LF is independent of geometry for a given overdensity.



(a) $\sigma = 4\,h^{-1}\mathrm{Mpc},\ \lambda_{\rm th} = 0.4$



(b) $\sigma = 10\,h^{-1}\mathrm{Mpc},\ \lambda_{\rm th} = 0.1$

Figure 4.5 The volume-limited LF and jackknife errors for each geometric environment, divided by the average LF of 9 realisations of shuffled catalogues composed of galaxies selected randomly from the full volume to mimic the density distribution of the corresponding geometric environment. Dashed lines represent a ratio of 1, indicating no variation between the geometric and the shuffled LFs. No statistically significant deviation away from a ratio of 1 is seen, which leads to the conclusion that the shuffled catalogues are consistent with the original LF and the cosmic web has no detectable direct influence.

where the density was defined by the smoothed density field (4 or $10\,h^{-1}$Mpc) used to initially generate the geometric classifications. This has the advantage that a volume limited sample is not required and hence it is possible to test the full magnitude range. Again no statistically significant difference was found, reinforcing the main result of this section, that the galaxy LF is independent of geometry for a given smoothed density.

4.1.3 Summary and discussion of LF results

The work detailed in the previous chapter, which developed a method for probing the cosmic web in galaxy redshift surveys which was then applied to the GAMA dataset, enabled a search for the impact of large-scale tidal forces on the galaxy population. There have been a number of studies suggesting that an effect of this sort should be present. Metuki et al. (2015) conducted a thorough analysis of galaxy properties within the cosmic web, finding significant dependencies on location, which they attributed largely to the strong relationship between galaxy properties and the properties of their host halos, which are in turn linked to their geometric environments. They found that the strong dependence of the halo mass function on the cosmic web was the main cause of the apparent dependence of galaxy properties. However, these analyses often do not test whether the relationships found can be directly attributed to a modulation by geometric environment, or rather are manifestations of the indirect influence of the local density field. In fact, Chapter 2 showed that, based on Gaussian statistics within the linear regime, we expect a variation of the halo mass function within different web components that is due solely to the underlying density field; there is no coupling to tidal forces and the theoretical halo mass function is independent of geometry at a given local density.

Yan et al. (2012) studied the tidal dependence of galaxy properties using the ellipticity, constructed from the eigenvalues of the tidal tensor in a way which exhibits less δ-dependence as a measure of environment than the classification method used in this work. In their analysis of SDSS data they found no physical influence of environmental morphology on galaxy properties. Similarly, Alpaslan et al. (2015, in prep.) investigated the relations between various galaxy properties and large scale structure identified within the GAMA dataset by the methods discussed in Section 3.3.2. They found the environment had limited direct impact

on galaxy properties, with the stellar mass of a galaxy playing a far larger role in shaping its evolution than the galaxy's location within the cosmic web. However, when identifying filaments by the 'Bisous' process, Guo et al. (2014) find a disparity between the LFs of satellite galaxies in SDSS whose host galaxy resides in a filament and those whose host galaxy does not, which they claim cannot be attributed to an environmental bias. Recent work has found a direct relationship between LSS anisotropies and the cosmic web when considering tensor properties such as the spin of galaxies (Libeskind et al. 2012) and angular momentum of dark matter substructures (Dubois et al. 2014). These authors find a correlation between the orientation of LSS and the axes of the cosmic web.

The main results of the GAMA-based LF analysis can be summarised as follows:

- The galaxy luminosity function in each component of the web was measured and a strong variation was seen. By fitting a Schechter function to each conditional luminosity function the variation of the LF between web components was quantified. The normalisation, described by ϕ^*, increases by a factor of $\sim$10 from voids to knots. The knee of the LF, M^*, brightens from voids to knots by 0.7 and 0.4 mag for 4 and $10\,h^{-1}$Mpc smoothing scales respectively. No clear trend in α, the parameter describing the faint end slope of the LF was seen.

- The direct influence of the cosmic web was tested by investigating the extent to which the observed modulation may be attributed to variations in the local density. By measuring densities over a range of scales between 6 and $12\,h^{-1}$Mpc it was shown that, in all cases, the modulation may be entirely accounted for by the variations in density between the geometric environments, indicating that *the galaxy LF is independent of geometry at a given local density.*

These results are consistent with the picture that scalar properties of halo and galaxy populations are independent of geometry – whilst a clear variation in the luminosity function of galaxies in different regions of the cosmic web was found, it was shown that this can be entirely attributed to the effect of the local density, reinforcing the idea that the underlying density field is the dominant property causing the relationship. The investigation as to whether the anisotropy

within different geometric environments is the cause of observed dependencies, or whether they can be fully accounted for by the distribution of densities alone is an important one in order to understand how to interpret new results in the context of the cosmic web. Whilst only the galaxy luminosity function has been considered thus far, there are a number of other observable properties that would benefit from a thorough analysis of the direct influence of the cosmic web. For example, the group mass function, as studied in the next section.

4.2 Group mass function and geometric environment

Chapter 2 investigated the dependence of the halo mass function on geometric environment within the MDR1 simulation. With the observed cosmic web classified within the GAMA survey, it is now possible to test the main conclusion of Chapter 2, that the halo mass function is independent of geometry for a given local density, by studying the modification of the observed group mass function by the geometric environments identified within GAMA. The same sample, $0.04 < z < 0.26$, selected in Section 3.1.1 is used throughout the remainder of this chapter.

Unfortunately, the limitations associated with observational catalogues make comparisons with theoretical predictions, such as those introduced in Chapter 2, nontrivial. The identification of groups, and the estimation of their masses, is significantly more complicated and uncertain than when using simulated data which track the underlying dark matter itself. Additionally, the predicted mass function will have a strong redshift dependence, and unlike the single redshift snapshot of the MDR1 simulation used in Chapter 2, the observed group catalogue covers a non-negligible redshift range and a smaller volume. This section first examines the global GAMA group mass function, its uncertainties and comparisons to theoretical models. Following this, the modulation of the group mass function by geometric environment is investigated.

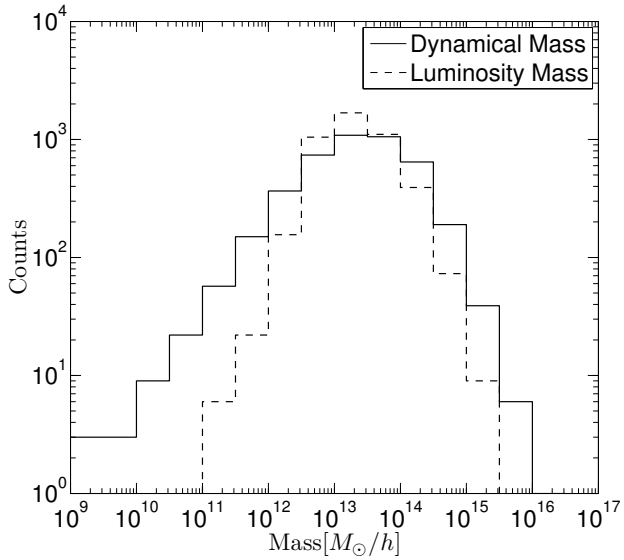


Figure 4.6 Comparison between the dynamical mass estimator and the luminosity mass estimator of GAMA groups, from Han et al. (2015).

4.2.1 GAMA global group mass function

As discussed in Section 3.1.3, there are two group mass estimates available from the GAMA group catalogue. The first is the dynamical mass, $M_{\rm dyn}$, estimated from the $M_{\rm dyn} \propto \sigma^2 R$ expected scaling of the group mass with the velocity dispersion, σ, and radius, R. Here the radius is defined as that which contains 50% of the galaxies in the group (see Robotham et al. 2011 for full details of the dynamical mass estimation). The second is the luminosity mass, $M_{\rm lum}$, which can be estimated using the expected proportionality of the group mass and luminosity, $M_{\rm lum} \propto L$. The dynamical masses are used to calibrate the luminosity-mass relation so that a simple scaling of the luminosity can provide a group mass estimate. The two estimates were discussed and compared by Han et al. (2015) (hereafter H15), as shown in Fig. 4.6. The dynamical mass can be seen to have a broader distribution and larger dispersion than the luminosity mass.

H15 compared the GAMA group mass estimates with weak lensing masses, finding the weak lensing mass measurements to be smaller than both the GAMA group mass estimates, though they note that the agreement with the weak lensing mass is slightly better with the luminosity mass than with the dynamical mass. It should be noted that the definition of a group mass is itself not well defined. It is the total mass within the group which is usually desired, and this is most

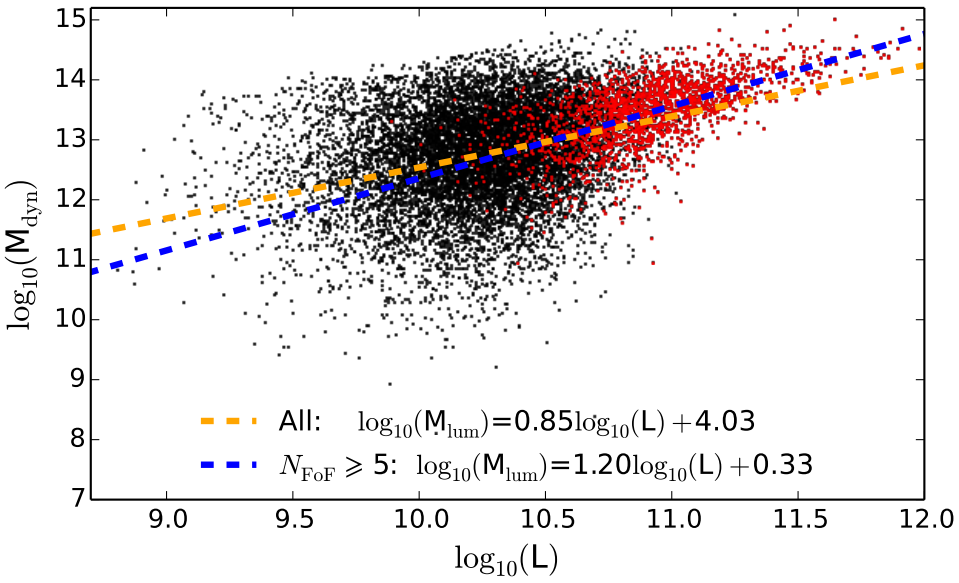


Figure 4.7 The dynamical mass and luminosity of all GAMA groups in the sample. Red points show groups with at least five members. The orange dashed line shows a best fitting linear (in $\log_{10}$) function to the two properties over all groups in the sample, the blue dashed line shows the fit over only those groups with at least five members.

often defined as the mass inside the viral radius. However, the virial radius is not clearly defined. Most often it is defined such that the average matter density inside the virial radius equals 200 times the mean density of the Universe. This definition of radius is implemented in the mass estimates of H15, which may be expected to produce different radii to those defined during the dynamical mass estimation.

H15 consider power-law combinations of six physical observables to find the best group mass estimator when matching to the weak lensing masses. They favour the construction $M \propto L^{\alpha}$ for its minimised scatter. Hence the construction

$$M_{\rm WL} = M_p \left(\frac{L}{L_0} \right)^{\alpha} \tag{4.4}$$

is introduced as a third mass estimator here, with $(\log_{10}(M_p), L_0, \alpha) = (13.51 \pm 0.13, 2 \times 10^{11} h^{-2} L_{\odot}, 1.09 \pm 0.22)$. Note that the subscript $_{\rm WL}$ here refers to the mass estimates calibrated from the weak lensing masses, not the weak lensing masses themselves.

Fig. 4.7 shows the scatter within the dynamical mass-luminosity plane for all

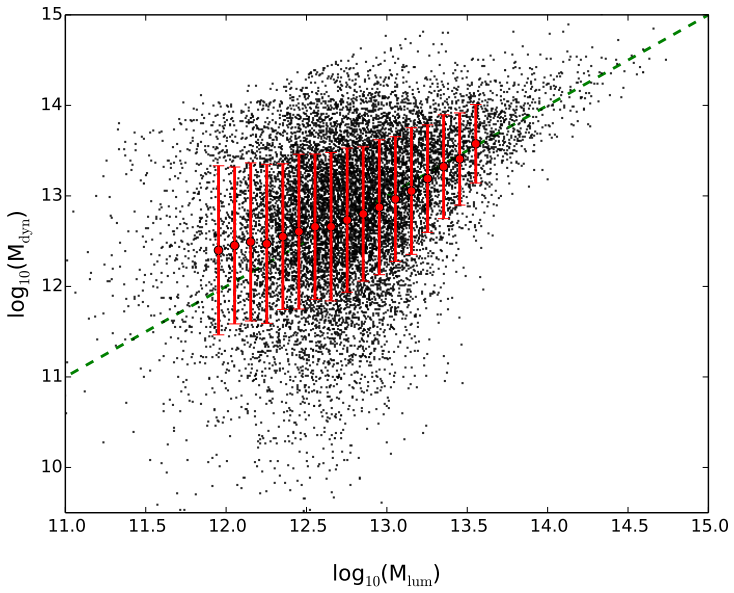


Figure 4.8 The dynamical, M_{dyn}, and luminosity, M_{lum}, mass estimates of the GAMA groups. Error bars show one standard deviation in $\log_{10}$ dynamical masses within bins of $\log_{10}$ luminosity mass. The average standard deviation over the bins is 0.74. The green dashed line plots $y = x$.

GAMA groups; the red points highlight those groups with at least five members. A least-squares fit to the dynamical mass-luminosity relationship ($\log_{10}(M_{dyn})$-$\log_{10}(L)$) of all groups in the sample is shown by the orange dashed line. A similar fit over only those groups with at least five members, motivated by the more accurate dynamical mass measurements of larger groups, is shown by the blue dashed line. The later fit is used to fix the parameterisation of the luminosity mass estimate as

$$\log_{10}(M_{lum}) = (1.20 \pm 0.03) \times \log_{10}(L) + (0.33 \pm 0.38). \quad (4.5)$$

Fig. 4.8 plots the resulting luminosity masses against the dynamical masses and it can be seen that there is significant disagreement between the two estimates for the majority of groups. This dispersion is quantified by calculating the standard deviation in $\log_{10}(M_{dyn})$ within bins of $\log_{10}(M_{lum})$, as shown by the red error bars in the figure. The average standard deviation in $\log_{10}(M_{dyn})$ is 0.74. Similarly, Fig. 4.9 plots the weak-lensing calibrated mass estimates against the dynamical masses and illustrates the scatter away from proportionality between the two properties. It can be seen that, as found by H15, the dynamical masses tend

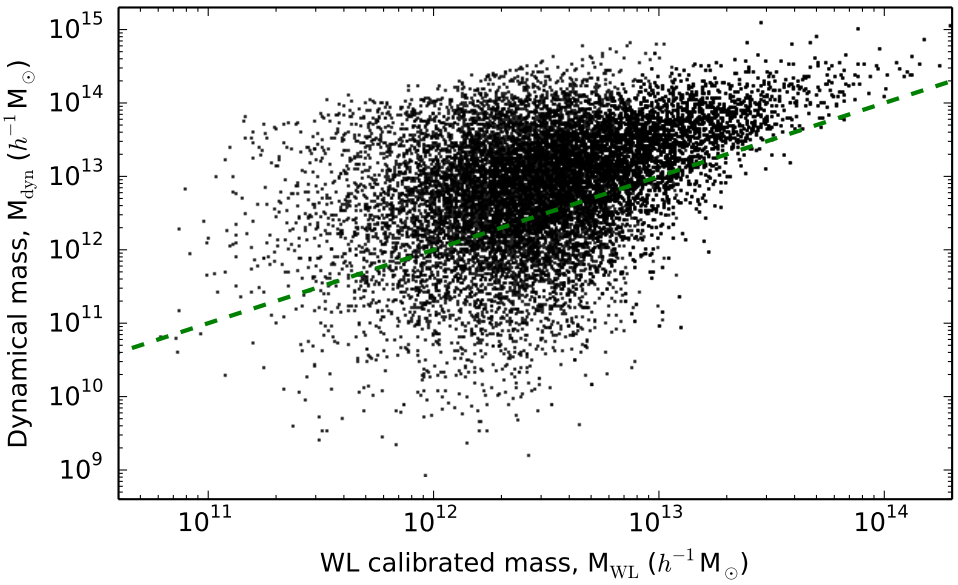


Figure 4.9 The dynamical masses and masses estimated from the luminosity using Eq. 4.4, $M_{\rm WL}$, calibrated using weak lensing masses. The green dashed line plots $y = x$, emphasising that the dynamical mass estimates are on average significantly larger than the weak-lensing-calibrated masses.

to be significantly larger than the M_{WL} masses estimated from Eq. 4.4. These figures show that there is a significant degree of uncertainty in the mass estimates of the GAMA groups.

A theoretical prediction for the mass function can be calculated in a similar manner to the Press-Schechter formalism discussed in Section 1.2.2. As in Chapter 2, the predictions calculated here are based on rescaling the universal collapsed fraction, which was estimated using the fitting formula proposed by Peacock (2007). The resulting multiplicity functions, F(M), a rescaled version of the mass function showing the fraction of mass in the Universe carried by objects in a unit range of ln(M), are shown in Fig. 4.10 for a range of redshifts. The red line in the figure shows the multiplicity function calculated from the $z = 0.1$ snapshot of the MDR1 simulation introduced in Chapter 2. It can be seen that the analytic prediction closely matches the simulated multiplicity function for all masses above the resolution limit of the simulation. At low masses below this resolution limit the simulation is not able to identify the halos and the multiplicity function drops off significantly. A similar lower mass limit will be present in the observational group catalogues; low mass groups cannot be consistently and

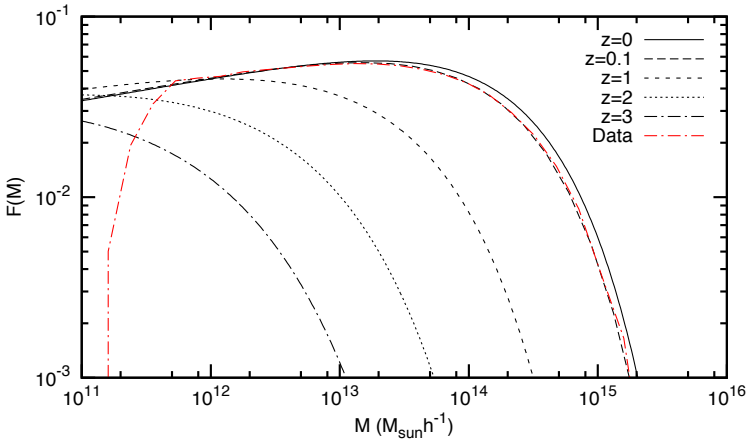


Figure 4.10 Black: The analytic predictions for the multiplicity function at different redshifts. **Red:** The multiplicity function calculated from the $z = 0.1$ snapshot of the MDR1 simulation introduced in Chapter 2.

accurately identified and we cannot expect the observed mass function to match the theoretical one at the low mass end. This figure also illustrates the magnitude of the redshift dependence of the theoretical multiplicity function.

Fig. 4.11 shows the group multiplicity function measured over the GAMA sample for the three mass estimates, $\mathrm{M_{dyn}}, \mathrm{M_{lum}}$ and $\mathrm{M_{WL}}$. It can be seen that there is significant disagreement between the three mass estimates. The shaded regions around the two scaled luminosity estimates, $\mathrm{M_{lum}}$ and $\mathrm{M_{WL}}$, represent one standard deviation uncertainty in the scaling parameters. $\mathrm{M_{lum}}$ is calibrated from the dynamical masses for only those groups with at least five members, hence, the uncertainty in the calibration for low mass groups is not included in the calculation of the uncertainty in the $\mathrm{M_{lum}}$ scaling. As found by H15, the luminosity mass results more closely match the weak lensing calibrated mass results than the dynamical masses. The large variation and uncertainty in these mass estimates presents a significant challenge for this analysis, which hopes to investigate potentially small variations in the mass function due to modulation by geometric environments. To illustrate the effect of these mass uncertainties on the resulting multiplicity function the theoretical mass function, $f(\mathrm{M})$, was smoothed with a Gaussian of varying widths along the $\log_{10}(\mathrm{M})$ axis to simulate mass uncertainties. More specifically, in order to ensure constant percentage error in mass, $Mf(\mathrm{M}) \equiv \mathrm{d}n/\mathrm{d}\ln(\mathrm{M})$ was convolved with a Gaussian. The broken green lines in Fig. 4.11 show the theoretical multiplicity function after smoothing with

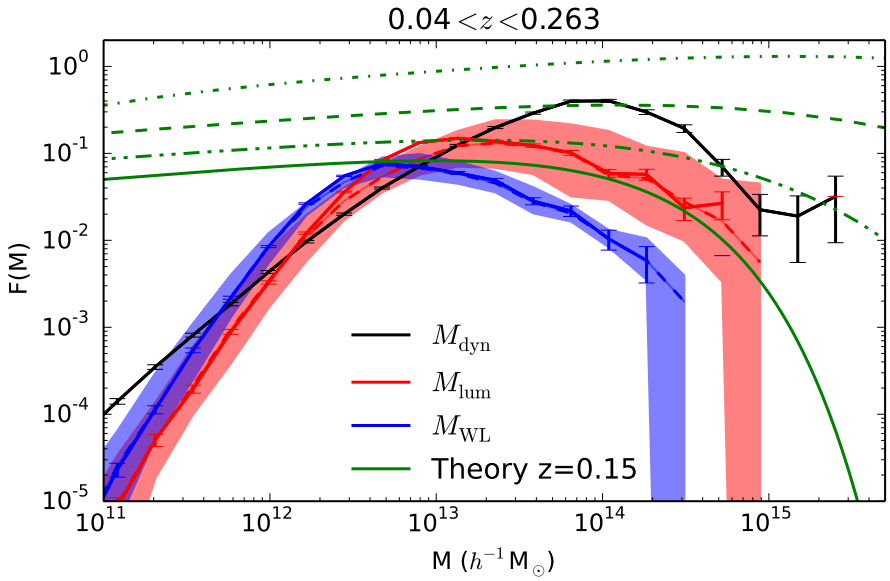


Figure 4.11 GAMA group multiplicity function for $0.04 < z < 0.263$, calculated for the three mass estimates as described in the text. For the dynamical mass measurement, error bars show Poisson errors on the number count in each bin. The shaded areas around the two scaled luminosity estimates, $\mathrm{M_{lum}}$ and $\mathrm{M_{WL}}$, represent one standard deviation uncertainty in the scaling parameters. The green solid line shows the theoretical multiplicity function at $z = 0.15$, the dash-dot-dot, dashed, and dot-dashed green lines show the theoretical multiplicity function after convolving $\mathrm{M}f(\mathrm{M}) \equiv \mathrm{d}n/\mathrm{d}\ln(\mathrm{M})$ with a Gaussian of width $\sigma = 0.5, 0.74$ and 1.0 along the $\log_{10}(\mathrm{M})$ axis respectively, to simulate mass errors of varying magnitudes. The significant divergence of the theoretical and observed multiplicity functions at the low mass end is due to heavy incompleteness at low masses.

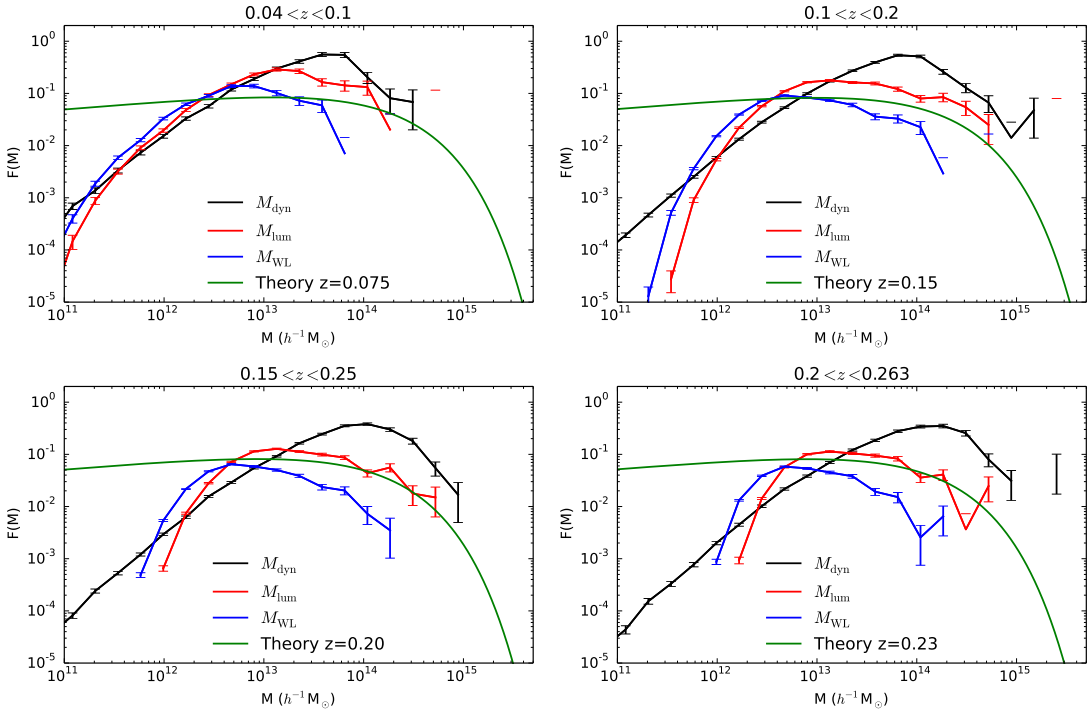


Figure 4.12 GAMA group multiplicity function, calculated over the redshift ranges given above each panel, for the three mass estimates as described in the text. Error bars show Poisson errors on the number count in each bin. The theoretical multiplicity function calculated for the redshift given in the key is shown in green.

Gaussians of width 0.5, 0.74 and 1.0, with the second value being the average standard deviation in $\log_{10}(\mathrm{M_{dyn}})$ within bins of $\log_{10}(\mathrm{M_{lum}})$ as shown in Fig. 4.8. It can be seen that mass uncertainties in this range significantly alter the resulting multiplicity functions, and can account for the difference in the peaks of the luminosity- and dynamical-mass multiplicity functions. The multiplicity functions as calculated over a variety of restricted redshift ranges are shown in Fig. 4.12, along with theoretical predictions without mass uncertainties. It can be seen that the disagreement between the three mass estimates remains across all redshifts considered. The theoretical multiplicity function is a moderately good match to the luminosity mass estimator, over a restricted mass range only.

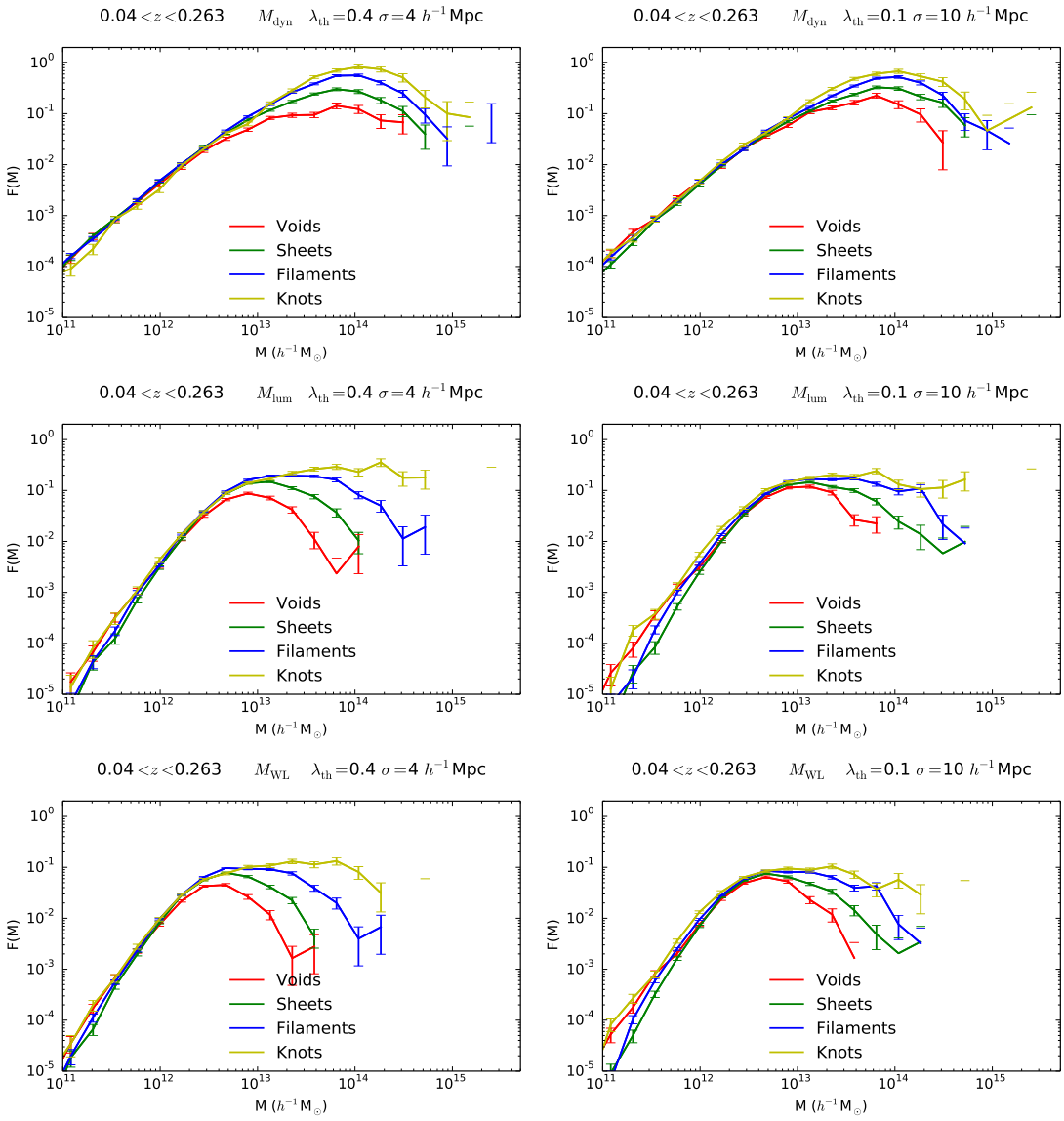


Figure 4.13 GAMA conditional group multiplicity functions, calculated over the redshift range $0.04 < z < 0.263$ for the three mass estimates; from top to bottom row the panels show dynamical masses, luminosity masses and weak-lensing-calibrated masses. Environments are classified with either $(\sigma, \lambda_{\rm th}) = (4\,h^{-1}{\rm Mpc}, 0.4)$ **(Left Column)** or $(10\,h^{-1}{\rm Mpc}, 0.1)$ **(Right Column)**. Error bars show Poisson errors on the number count in each bin.

4.2.2 GAMA group mass function within the cosmic web

The multiplicity function is defined as

$$F(M) = M^2 f(M)/\bar{\rho}. \tag{4.6}$$

Hence, a conditional multiplicity function dependent on geometric environment can be extracted by treating each component of the cosmic web as a separate sample, with its own individual average density, $\bar{\rho}_{\rm env}$. That is, we need to know the total volume, V, of a given component of the web as well as the total mass within it. Then, with this definition, the multiplicity function will be independent of V unless there is some dependence of halo masses on environment. The mass function of each environment, $f(M)_{\rm env}$, will be the comoving number density of objects within $V_{\rm env}$, the volume of the sample classified as each environment. Despite the uncertainty remaining in the global mass function, it is still fruitful to look at the conditional mass functions in order to compare the differences between each component of the cosmic web for a given mass estimate and investigate the modulation by geometric environment. Fig. 4.13 displays the conditional multiplicity functions for the geometric environments calculated with the two parameter sets chosen in Section 3.2.3, $(\sigma, \lambda_{\rm th}) = (4\,h^{-1}{\rm Mpc}, 0.4)$ and $(10\,h^{-1}{\rm Mpc}, 0.1)$. Each row of the figure plots the multiplicity functions calculated with one of the three mass estimates. It can be seen that in all cases, the magnitude of the function increases between voids and knots. The functions for each environment are very similar for low masses, where the group catalogues are heavily incomplete; close to the turning point of the multiplicity function the results for the four environments begin to diverge. For all three mass estimates, the results of the smaller smoothing scale parameter set produces larger disparity between the multiplicity functions of the environments.

The main conclusion of the Gaussian theory of Chapter 2 was that the only dependence of the halo multiplicity function on geometric environment is due to the different overdensities sampled by the environments. Hence, for a given overdensity, the distribution of halo masses would be predicted to be the same regardless of location within the cosmic web. As with the simulated halo mass function in Chapter 2, this can be tested directly within the GAMA group catalogue by comparing the multiplicity functions of groups selected from

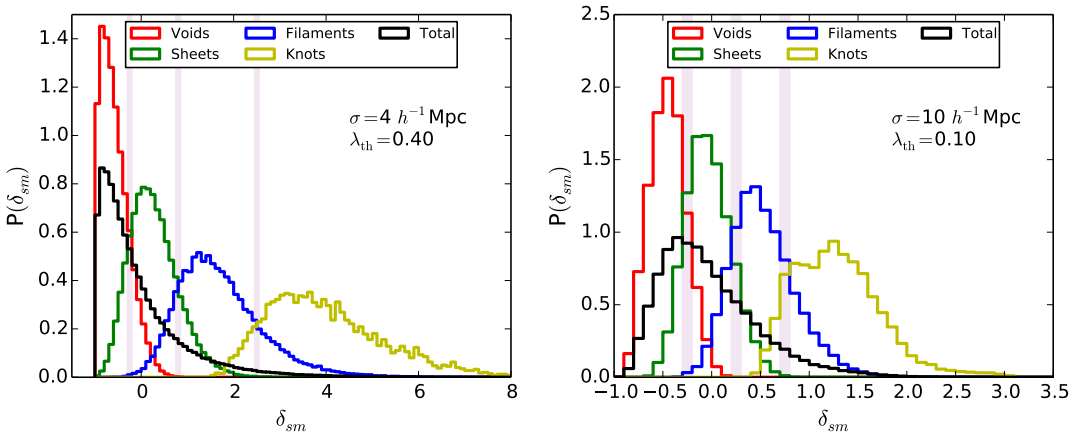


Figure 4.14 Distribution of cell overdensities, $\delta_{\rm sm}$, for each geometric environment, after smoothing with $\sigma = 4$ or $10\,h^{-1}$Mpc as shown in the key. The purple shaded regions indicate the overdensity ranges selected for the restricted density multiplicity functions.

environments within a specific density range. The distribution of overdensities within each environment are shown in Fig. 4.14. Here the overdensity, $\delta_{\rm sm}$, is the overdensity of each cell, measured as described in Section 3.2.1, after smoothing with a Gaussian of width $\sigma = 4$ or $10\,h^{-1}$Mpc. As expected, the larger smoothing reduces the magnitude of the average overdensities. Each environment from voids to knots samples, on average, more overdense regions, however there is significant overlap between the environments which allows the hypothesis to be tested in this manner. The purple shaded regions in Fig. 4.14 indicate the overdensity ranges ($\Delta\delta_{\rm sm} = 0.1$) selected for the comparison. Each range is selected in order to include a comparable number of cells from at least two geometric environments.

The conditional multiplicity functions of these restricted density regions are shown in Fig. 4.15, Fig. 4.16 and Fig. 4.17 for the dynamical, luminosity and weak-lensing-calibrated mass estimates respectively. For each of the mass estimates, smoothing scales and overdensity ranges there is a strong agreement between the multiplicity functions of the different geometric environments. This is further emphasised by the subpanels at the bottom of each panel which show the ratio of the multiplicity functions of each geometric environment to the total multiplicity function measured over all of the volume with a smoothed overdensity within the specified range. The dashed line in each subpanel shows a ratio of 1, indicating an equal conditional multiplicity function to the total multiplicity function. Statistically the results are consistent with the picture that the group mass function is independent of geometric environment for a given overdensity.

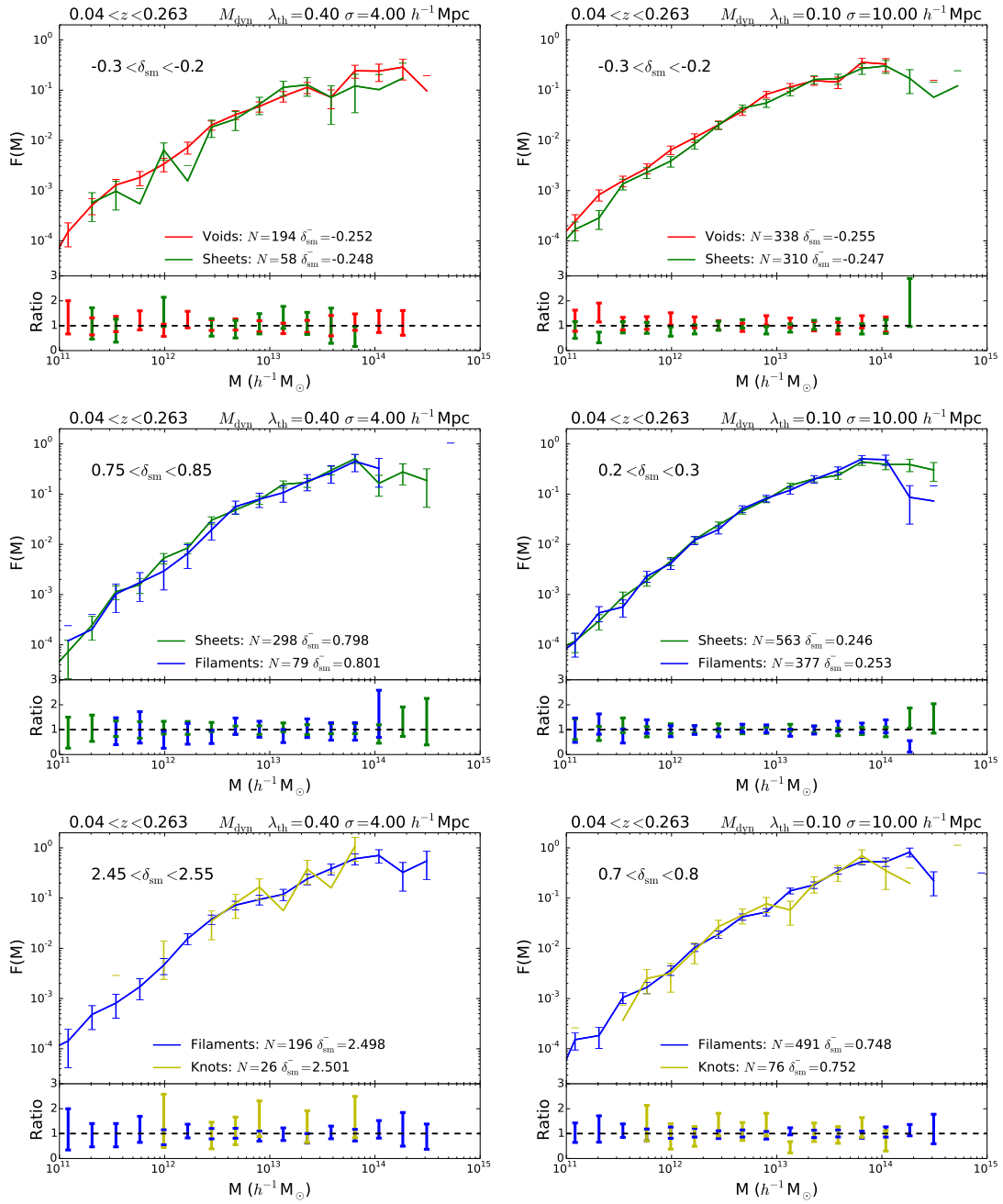


Figure 4.15 GAMA conditional group multiplicity functions, calculated over the redshift range $0.04 < z < 0.263$ for the dynamical mass estimates. Each conditional multiplicity function is measured only over those cells whose smoothed overdensity falls within a specific range, as stated in the top left of each panel. Environments are classified with either $(\sigma, \lambda_{\rm th}) = (4\,h^{-1}{\rm Mpc}, 0.4)$ **(Left Column)** or $(10\,h^{-1}{\rm Mpc}, 0.1)$ **(Right Column)**. Error bars show Poisson errors on the number count in each bin. The bottom subpanel shows the ratio of each conditional multiplicity function to the total multiplicity function measured over all cells within the specified overdensity range. For each environment, the number of groups found to lie within the specified density range, N, and the average overdensity of all cells within the specified range, $\bar{\delta}_{\rm sm}$, is shown in the key.

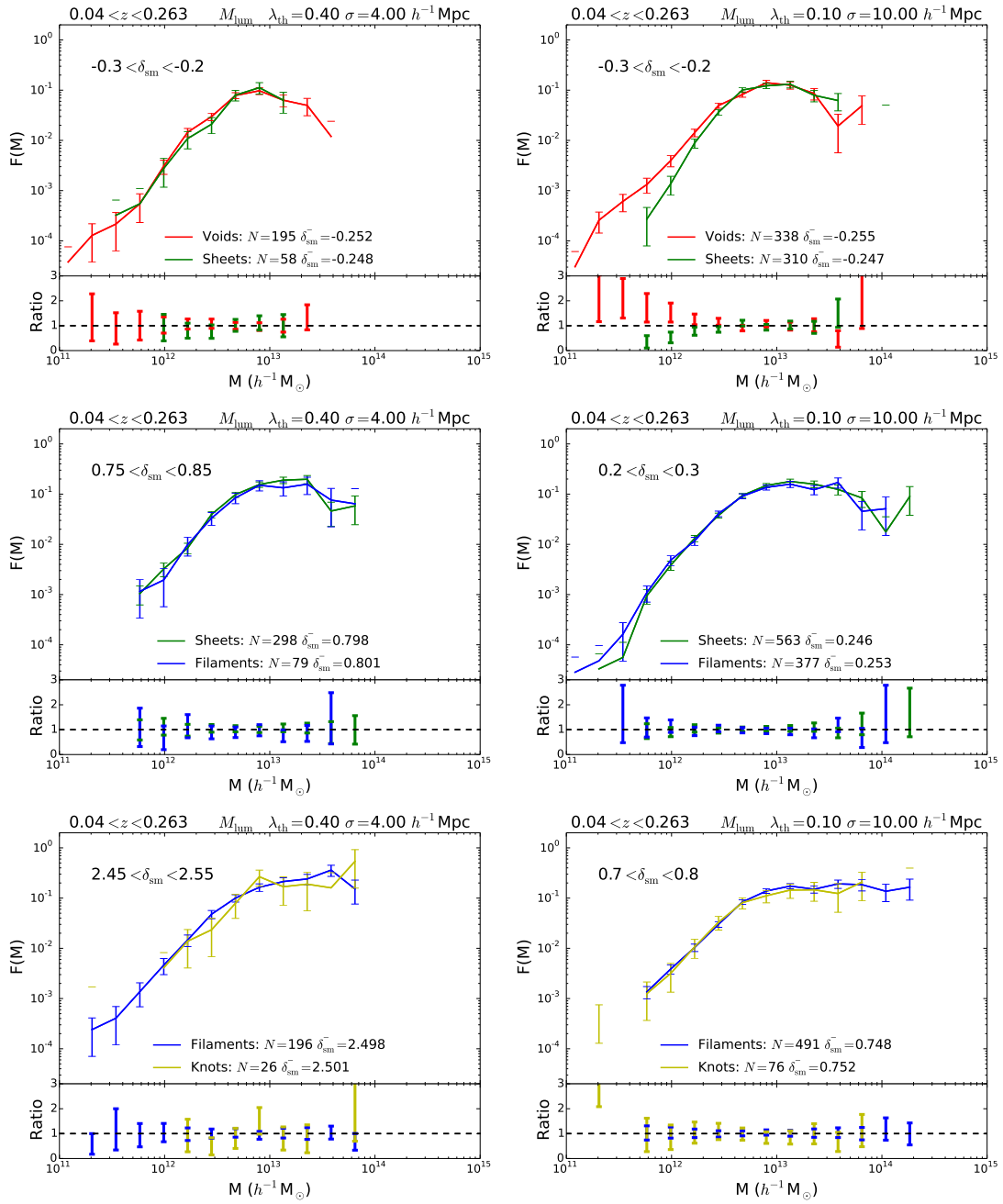


Figure 4.16 GAMA conditional group multiplicity functions, calculated over the redshift range $0.04 < z < 0.263$ for the luminosity mass estimates. Each conditional multiplicity function is measured only over those cells whose smoothed overdensity falls within a specific range, as stated in the top left of each panel. Environments are classified with either $(\sigma, \lambda_{\rm th}) = (4\,h^{-1}{\rm Mpc}, 0.4)$ **(Left Column)** or $(10\,h^{-1}{\rm Mpc}, 0.1)$ **(Right Column)**. Error bars show Poisson errors on the number count in each bin. The bottom subpanel shows the ratio of each conditional multiplicity function to the total multiplicity function measured over all cells within the specified overdensity range. For each environment, the number of groups found to lie within the specified density range, N, and the average overdensity of all cells within the specified range, $\bar{\delta}_{\rm sm}$, is shown in the key.

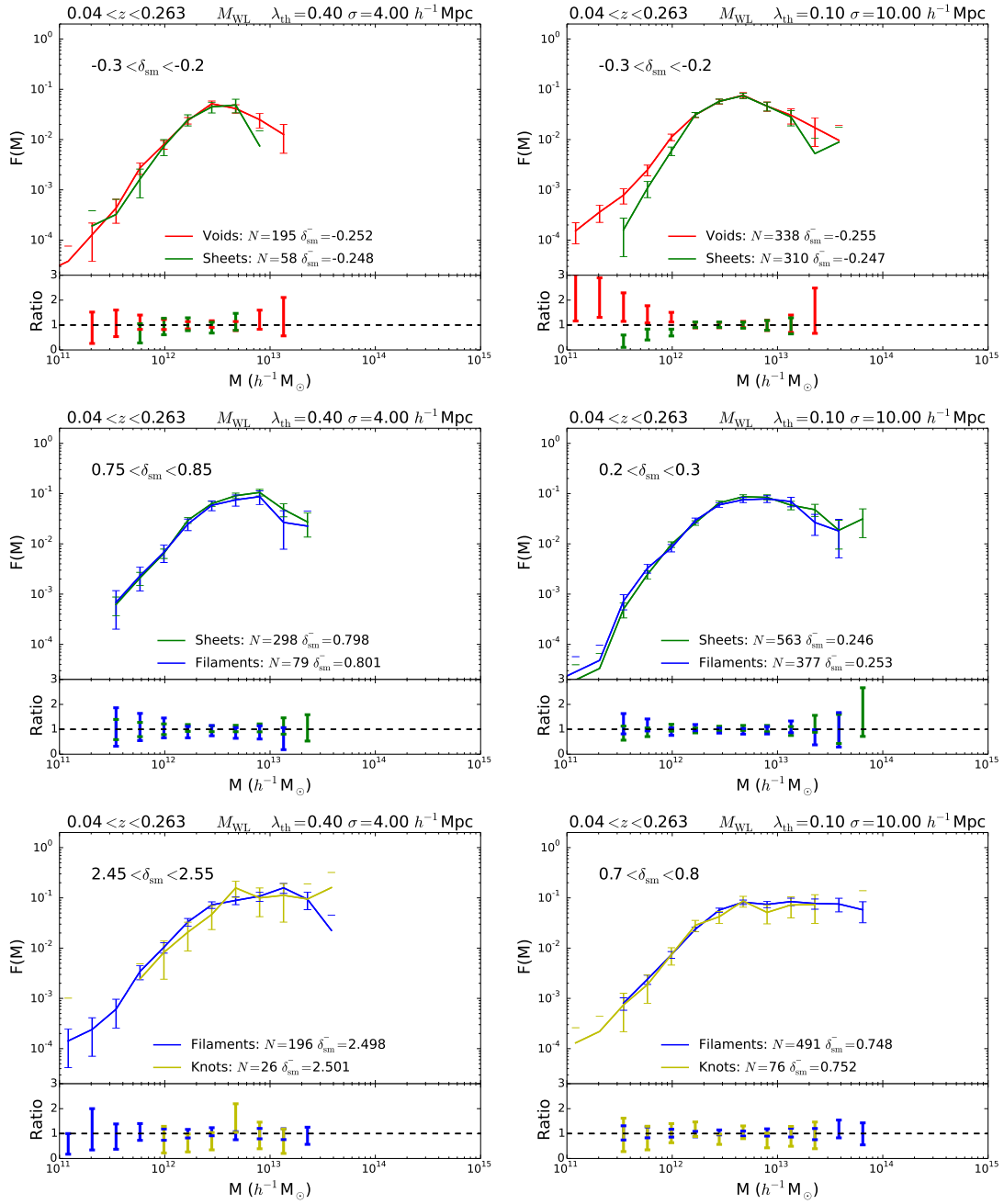


Figure 4.17 GAMA conditional group multiplicity functions, calculated over the redshift range $0.04 < z < 0.263$ for the weak-lensing-calibrated mass estimates. Each conditional multiplicity function is measured only over those cells whose smoothed overdensity falls within a specific range, as stated in the top left of each panel. Environments are classified with either $(\sigma, \lambda_{\rm th}) = (4\,h^{-1}{\rm Mpc}, 0.4)$ **(Left Column)** or $(10\,h^{-1}{\rm Mpc}, 0.1)$ **(Right Column)**. Error bars show Poisson errors on the number count in each bin. The bottom subpanel shows the ratio of each conditional multiplicity function to the total multiplicity function measured over all cells within the specified overdensity range. For each environment, the number of groups found to lie within the specified density range, N, and the average overdensity of all cells within the specified range, $\bar{\delta}_{\rm sm}$, is shown in the key.

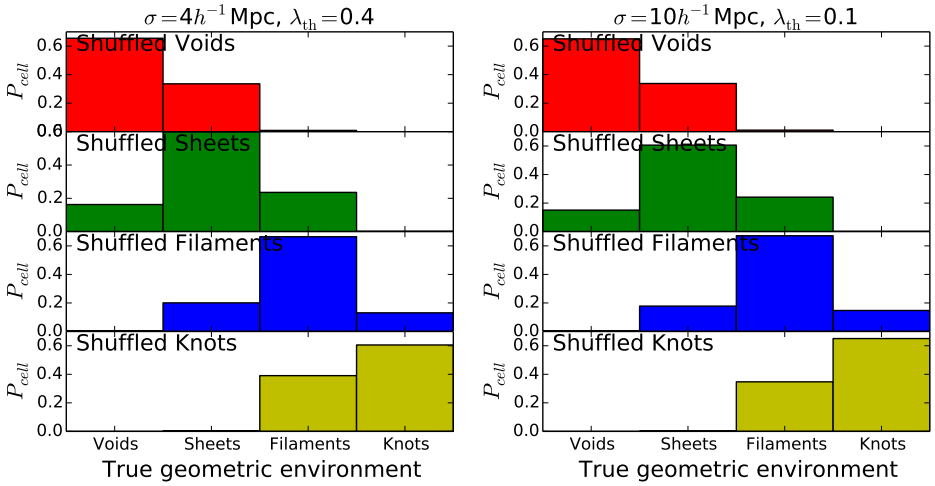


Figure 4.18 The proportion of cells from each geometric environment used to create the 'shuffled' environments, averaged over 20 realisations of shuffled catalogues. Each row represents one of the shuffled environments, the relative heights of the bars in each row illustrates the proportion of cells from each 'true' geometric environment in the shuffled environment. True geometric environments are classified with either $(\sigma, \lambda_{\rm th}) = (4\,h^{-1}{\rm Mpc}, 0.4)$ **(Left)** or $(10\,h^{-1}{\rm Mpc}, 0.1)$ **(Right)**.

The limited number of groups available within the GAMA dataset restricts the strength of this conclusion. There may be small modulations due to geometric environment which are not visible with this method due to the statistical uncertainties arising from the limited number of groups which fall within each environment, mass bin and density range. For that reason the environmental dependence is also investigated in a similar manner to Section 4.1.2. All cells within the sampled volume are reclassified by selecting at random a new 'shuffled' environment classification with a probability proportional to the relative abundance of 'true' geometric environments of all cells with the same overdensity as the cell in question. That is, a cell of overdensity $\delta_{\rm sm}$ will be assigned a new environment, e', with probability

$$p_{e'} = N_{e',\delta_{\rm sm}}/N_{tot,\delta_{\rm sm}}, \tag{4.7}$$

where $N_{e',\delta_{\rm sm}}$ is the number of cells of environment e' within the density bin containing $\delta_{\rm sm}$, and $N_{tot,\delta_{\rm sm}}$ is the total number of cells within that bin. Fig. 4.18 shows, for the cells which make up each shuffled environment, the distribution of their original environments before the reclassification process. Due to form of the overdensity distributions shown in Fig. 4.14, this shuffling procedure results

in approximately 60% of cells retaining their original environment classification. The resulting classifications have overdensity distributions identical to the distributions of the original environments, however, the geometric environment information has been removed by this shuffling process.

If the geometric environment introduces a modulation of the group mass function which cannot be accounted for by the difference in the densities between the environments, then the multiplicity functions of these shuffled environments should differ from the original multiplicity functions shown in Fig. 4.13. 20 shuffled multiplicity functions were created from 20 realisations of these shuffled environments by reclassifying the groups according to the new classification of the cell in which they reside. Fig. 4.19 compares the original and the shuffled group multiplicity functions. The conditional multiplicity functions calculated from the true geometric environments are shown in dashed coloured lines. The colour coded shaded regions indicate two standard deviations around the average conditional multiplicity functions of the 20 shuffled catalogues. It can be seen that, in general, the original multiplicity functions are consistent with the shuffled results. The ratios of the original and shuffled multiplicity functions are shown in Fig. 4.20. This figure emphasises the consistency with a ratio of 1, as shown by the dashed line in each panel, for the majority of environments and mass estimates, particularly around M$\approx 10^{13}\ h^{-1}\mathrm{M}_{\odot}$, which is the peak in group mass distribution as shown in Fig. 4.6. However, there are still deviations beyond the statistical expectations in the case of no direct dependence, although it should be noted that the mass uncertainties have not been taken into account when calculating the error bars. The low mass end, particularly for the larger smoothing scale, shows the strongest disparity, along with the groups found in void regions defined with the smaller smoothing scale. This analysis does not use a volume limited sample, hence some redshift dependence may remain. If the redshift distribution of the original environment is significantly different to the resulting distributions of the shuffled environment, the redshift dependence of successful identification of groups and the accuracy of their mass estimates may introduce additional uncertainty into the results. Overall, the substantial uncertainties in the global group mass function and the limited sample of groups available within GAMA limit the confidence with which any conclusions can be made and restrict the magnitude of modulation by geometric environment which can be identified

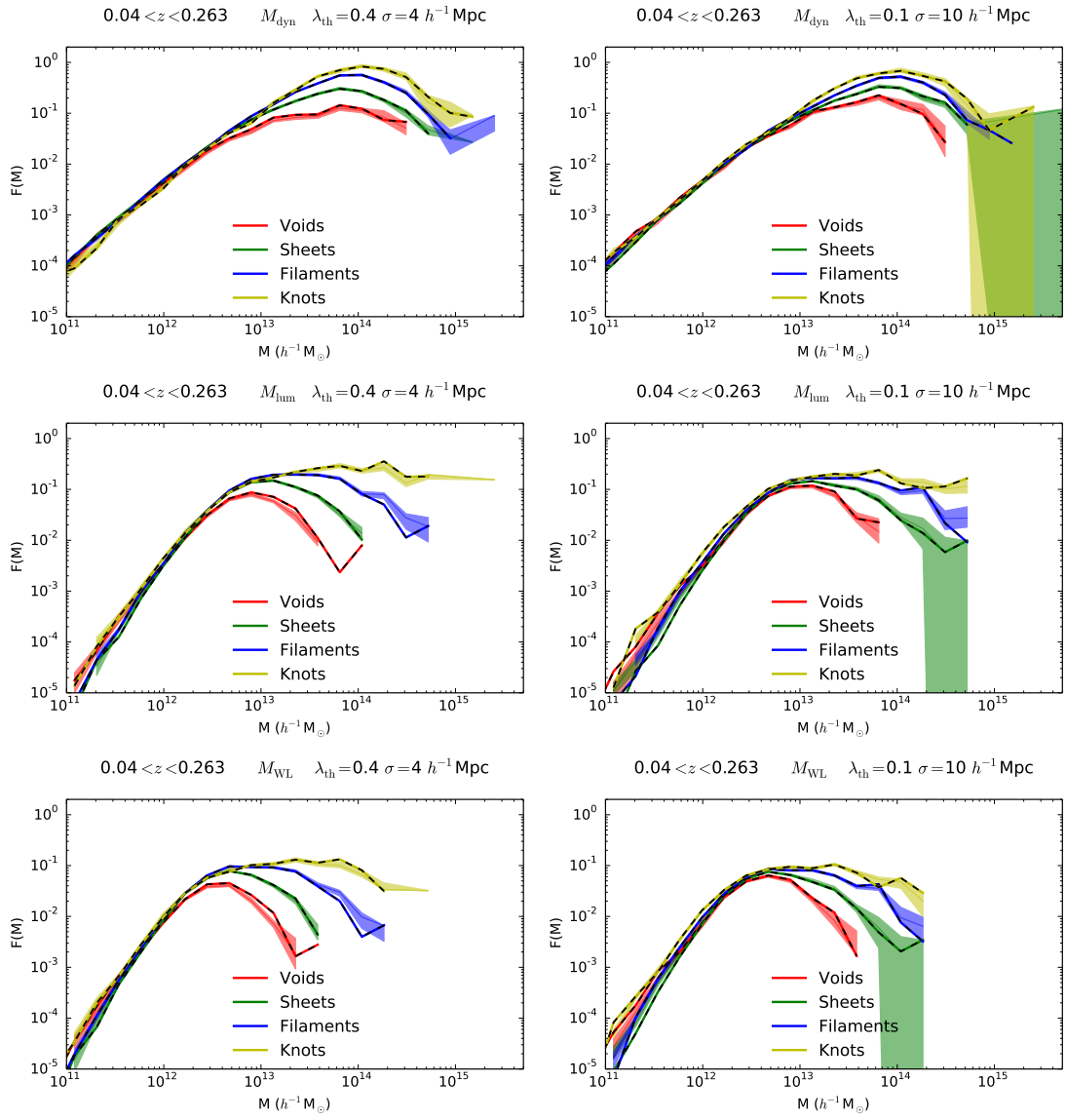


Figure 4.19 GAMA conditional group multiplicity functions, calculated over the redshift range $0.04 < z < 0.263$ for the three mass estimates; from top to bottom row the panels show dynamical masses, luminosity masses and weak-lensing-calibrated masses. Dashed lines colour coded as shown in the key plot the multiplicity functions of the original geometric environments, classified with either $(\sigma, \lambda_{\rm th}) = (4\,h^{-1}{\rm Mpc}, 0.4)$ **(Left Column)** or $(10\,h^{-1}{\rm Mpc}, 0.1)$ **(Right Column)**. Colour coded shaded regions indicate ± 2 standard deviations around the average of 20 shuffled multiplicity functions, randomly classified in such as way as to reproduce the overdensity distributions of the original geometric environments.

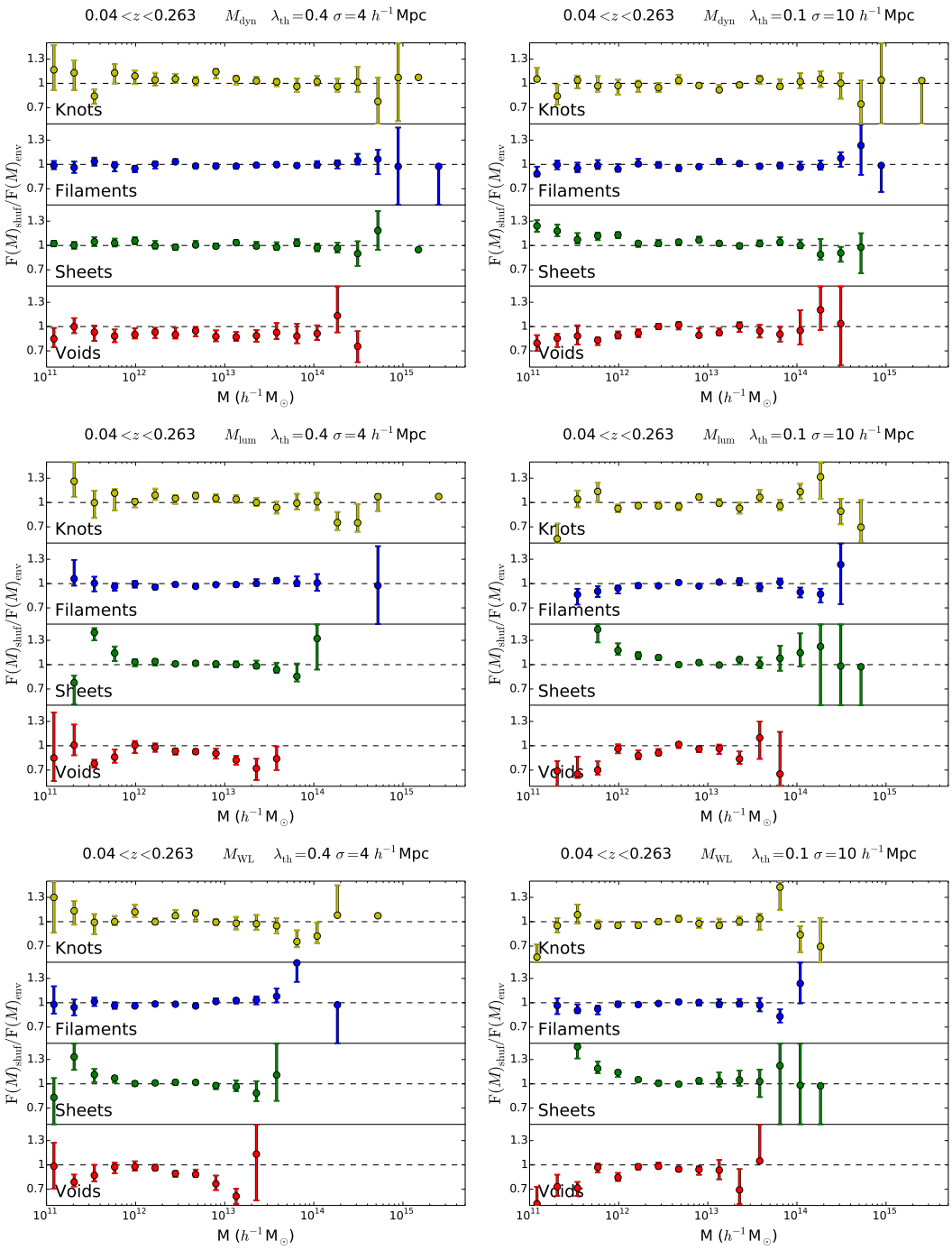


Figure 4.20 The average multiplicity function and ± 2 standard deviation errors of 20 realisations of shuffled catalogues divided by the multiplicity functions of the true geometric environments. From top to bottom row the panels show results using dynamical masses, luminosity masses and weak-lensing-calibrated masses. Shuffled catalogues are composed of cells reclassified randomly from the full volume to mimic the density distribution of the corresponding geometric environment. Dashed lines represent a ratio of 1, indicating no variation between the geometric and the shuffled MFs.

or ruled out. However, this section has found no strong evidence of a direct dependence of the GAMA group mass function on location within the cosmic web beyond the modulation which can be accounted for by the different densities sampled by the environments. The next chapter presents the final study of this thesis, in which the attention is turned to the *histories* of assembly, rather than the final state properties investigated thus far.

Chapter 5

Stellar assembly histories within the cosmic web

All investigations presented so far in this thesis have found no direct evidence of an impact of the tidal force aspect of the environment. That is, all investigations have been consistent with the picture that scalar properties of large-scale structure are independent of location within the cosmic web for a given local overdensity. However, the scalar properties considered, the galaxy luminosity and group and halo mass functions, all look at the 'final state' of objects. It is conceivable that whilst no impact of tidal forces has been detected when looking at final states, the *history* of these objects may depend more directly on their location within the cosmic web.

The spatial distribution and clustering of dark-matter halos (Mo & White 1996, Sheth & Tormen 1999, Cooray & Sheth 2002), as well as the connection between galaxies and the halos that host them (Berlind & Weinberg 2002, Magliocchetti & Porciani 2003, Contreras et al. 2015) have been extensively studied. Theoretical models, such as the excursion set formalism discussed in Chapter 2, support the expectation that halo mass is the dominant property influencing halo abundance and clustering (Press & Schechter 1974, Bond et al. 1991). Whilst this has been verified in simulations by some studies (Lemson & Kauffmann 1999, Percival et al. 2003), other studies have shown that mass is not the only property upon which halo clustering depends; the term 'halo assembly bias' is used to refer to the dependence of halo clustering on properties other than halo mass. Recent studies have detected halo assembly bias in simulations

and have found that halos of the same mass cluster differently depending on how long ago they assembled their mass (Gao et al. 2005, Wechsler et al. 2006, Gao & White 2007). Many halo properties, such as concentration and spin, exhibit a dependence on halo formation time, and hence, would also be expected to exhibit assembly bias (Faltenbacher & White 2010). A correlation of galaxy properties with these history-dependent halo properties would contradict the simple galaxy-halo relationships which assume a dependence on mass alone and induce systematic errors (Zentner et al. 2014). However, results of observational investigations of assembly bias are less clear, with some studies claiming to have detected assembly bias (Yang et al. 2006, Wang et al. 2013), and others finding no evidence of assembly bias (Blanton & Berlind 2007, Tinker et al. 2008).

A similar importance is placed on halo mass within theoretical models of the galaxy-halo connection. Standard formalisms used to predict the galaxy occupation statistics of halos, such as the Halo Occupation Distribution (HOD, Peacock & Smith 2000, Berlind et al. 2003, Zheng et al. 2005) and the Conditional Luminosity Function (Yang et al. 2003, van den Bosch et al. 2007), assume a dependence on halo mass alone. Whilst these models have had much success, there is reason to question, or at least a need to verify, the simplicity of such models. Zentner et al. (2014) compared the HODs recovered from mock galaxy catalogues with and without assembly bias, finding that the galaxy-halo relationship inferred from galaxy clustering is subject to significant systematic errors induced by assembly bias. Hence, a better understanding of the extent to which assembly bias exists, and how it affects the galaxy-halo relationship is of high importance.

Continuing the investigation into how the cosmic web influences galaxies and halos, this chapter studies assembly histories of stellar populations within different components of the cosmic web. Any dependence of stellar-mass assembly on geometric environment for fixed group mass would provide a strong indication of assembly bias. In turn, such results would allow us to improve the standard halo occupation statistic formalisms and provide a better way to re-parameterise galaxy-halo relation models, whilst also improving our understanding of the connection between collapsed objects and their large-scale environments.

This chapter is structured as follows: the recovery of star formation histories from the GAMA spectra, by implementation of the VESPA algorithm, is discussed

in Section 5.1. The optimal estimator of assembly time is considered in Section 5.2 where the global star formation histories are presented. Section 5.3 investigates how assembly time varies within the cosmic web of large-scale structure and Section 5.4 summaries and discusses the results of this chapter.

5.1 Measuring star-formation histories

The spectrum of light emitted by a galaxy contains a wealth of information about its history, components and physics. Emission lines indicate recent star formation, the intensity of which provides information on the abundance of young stars. Absorption features are related to chemical abundances, and trace stellar populations, their ages and metallicities. The overall shape of the spectrum can tell us about star-formation history and dust. A galaxy's spectrum is simply the combined spectra of all the individual stars and other radiating matter within the galaxy, and hence it can be modelled as a superposition of stellar populations of different ages. Simple Stellar Population models (SSPs) use an assumed initial mass function, stellar evolution and spectral libraries to model the spectra emitted at time t by a population of stars created at an earlier time t_0. Using these SSPs, a galaxy's spectrum can be described via superposition in terms of its star-formation history. However, it should be noted that such a technique is heavily reliant on the models and is therefore limited by the accuracy with which one can model stellar populations. The post-main-sequence phases of stellar evolution are particularly complex, so different codes may yield systematically different results. This is less of an issue for the current work, which is concerned with detecting differences in the results as a function of environment.

5.1.1 The VESPA algorithm

VESPA (VErsatile SPectral Analysis) is a code to recover star formation and metallicity histories, dust content and present-day stellar masses from galactic spectra, by splitting each galaxy into a sum of simple stellar populations. The full details of the algorithm are presented in Tojeiro et al. (2007). For the application of VESPA to the GAMA spectra, the SSP models of Maraston (2005) are used. The modelling of dust extinction employs a one-parameter dust model and a single dust screen is applied to the combined luminosity of all the galactic

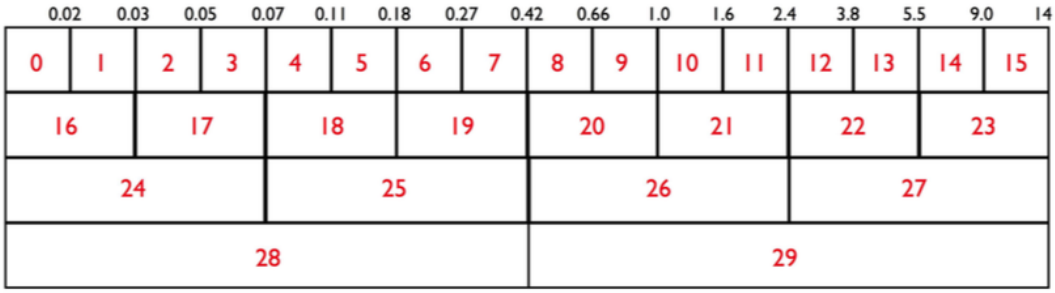


Figure 5.1 Schematic view of the grid of bins used by VESPA. The top line of black numbers indicates the age of each boundary, in Gyrs. The red number in each of the bins is a unique bin identifier number, which can be used to quickly retrieve the properties of a given bin. (From Tojeiro et al. 2009)

components.

As mentioned above, inaccuracies in the modelling and noise in the data can introduce degeneracies and uncertainties into the results. Therefore one must limit the amount of the information attempted to be extracted to avoid the parameter space becoming unsustainably large. The unique appeal of VESPA is that, motivated by the realisation that the amount of information which can be safely recovered from galactic spectra varies on a galaxy-by-galaxy basis, the number of parameters VESPA attempts to recover for each galaxy is allowed to adapt. VESPA recovers stellar mass fractions across age bins of varying sizes, as shown in Fig. 5.1. For each galaxy VESPA estimates the resolution with which the history can be estimated and provides results for the appropriate sized bins accordingly.

5.1.2 The GAMA spectroscopic data

Data from the GAMA spectroscopic redshift survey, discussed in Section 3.1 is used (see Hopkins et al. (2013) for full details of the spectroscopic analysis). The spectra of any spectroscopic survey require flux calibration, to correct the wavelength dependence of the system throughput, and to provide an approximate absolute flux calibration. The GAMA spectra display imperfect flux calibration; Hopkins et al. (2013) find the GAMA flux calibration to be typically accurate to $10-20\%$, with poorer reliability at the extreme wavelength ends, and poorer in the blue than the red. The flux calibration was tested by comparison with the SDSS fibre spectroscopy. 574 objects which were observed by both GAMA and

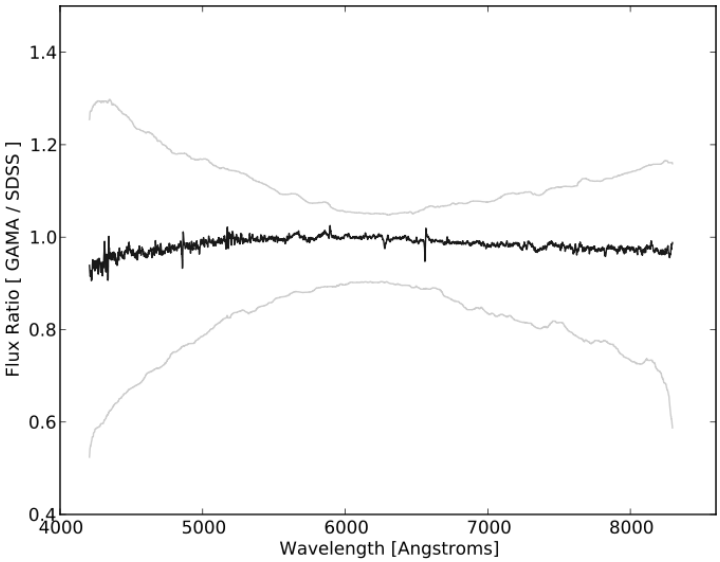


Figure 5.2 The median ratio of 574 common GAMA and SDSS spectra. The spectra are first normalised by the median flux value of the flux-calibrated spectrum, (since the absolute flux calibration is scaled differently between GAMA and SDSS). The spectra are then median filtered by 7Å and interpolated to the same wavelength scale before the ratio was taken. The dark line shows the median of the flux ratios, and the outer, grey lines show the 68th percentile range of the distribution of ratios for individual objects. From Hopkins et al. (2013).

SDSS were compared. Fig. 5.2 shows the median and 68th percentile of the ratio of the two sets of flux-calibrated spectra, after normalisation to the median flux of each spectrum. It can be seen that the agreement is best in the mid-wavelength range, but noisier at the extreme blue and red ends of the spectrum. Hopkins et al. (2013) note that there remains an unresolved issue with the level of the response in the blue end of the spectra.

Unlike single-line fitting, any full spectral-fitting code such as VESPA is sensitive to the overall shape of the continuum. For this reason, VESPA was adapted to also take photometric information into account in addition to the spectroscopic data. Where the photometric bands overlap the observed spectra, they can be used to calibrate the spectrum by comparing with the result of synthetic photometry. Otherwise, VESPA can be required to predict broad-band magnitudes and use these as part of its fitting. The addition of u-band data in this way was especially important. The GAMA flux calibration includes a final step where the spectrum is scaled to the r-band Petrosian magnitude, but an analysis of the match between the individual spectra and synthetic photometry

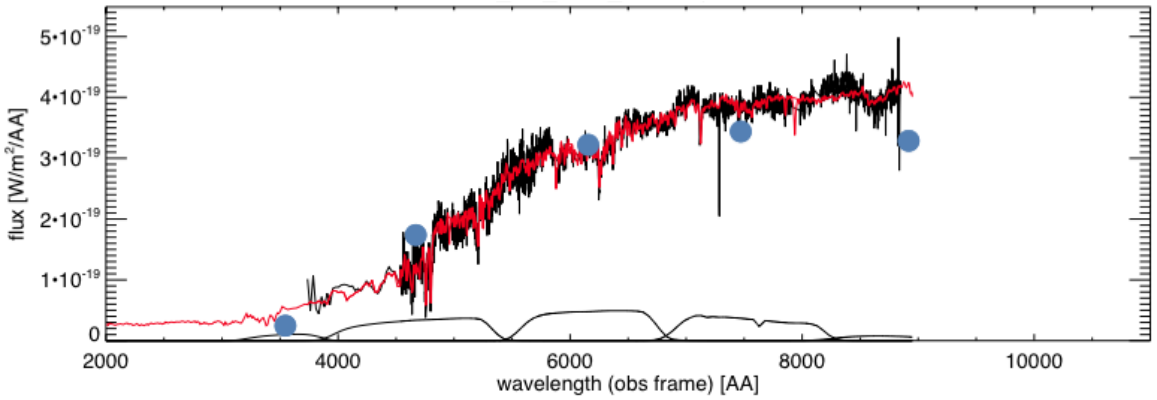
indicated that the calibration was imperfect and the spectra required additional calibration before application of VESPA.

The photometry used was the aperture-matched Petrosian magnitudes from SDSS. Petrosian magnitudes attempt to measure a constant fraction of the total light, independent of the position and distance of the object (Petrosian 1976). Hence the Petrosian system measures galaxy fluxes within a circular aperture whose radius is defined by the shape of the azimuthally averaged light profile. The GAMA spectra cover the g, r and i photometry bands; it is desirable that the integrated spectra across the wavelengths of each individual band matches the photometric flux.

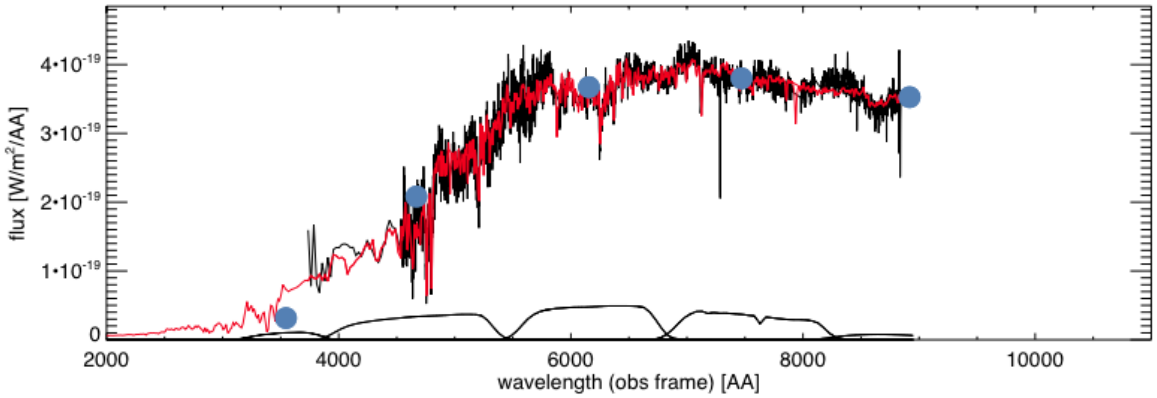
The SDSS filter transmission functions were interpolated to the wavelengths of the spectra to provide weightings for the integration, allowing for an estimate of the flux that would contribute to the photometry of each band. Comparison with a 'standard spectrum' of constant flux whose expected magnitude can be calculated allows for an estimate of the g, r and i magnitudes of each spectrum. The difference between these 'spectro-magnitudes' and the SDSS magnitudes provides an estimate for the flux calibration error at the effective wavelength of each filter. A linear interpolation was implemented in two regimes, across the blue and red side of the r-band effective wavelength individually, to provide an estimated magnitude difference for each wavelength of the spectrum, $\Delta\mathrm{M}(\lambda)$. At each wavelength, the flux was scaled by the appropriate linearly-interpolated value, $k(\lambda)$, once the magnitude-quantity had been converted to a flux-quantity using the equation

$$k(\lambda) = F_{\mathrm{scaled}}(\lambda)/F_{\mathrm{original}}(\lambda) = 10^{0.4\times\Delta\mathrm{M}(\lambda)}. \tag{5.1}$$

This process was repeated three times for each spectrum, where each iteration used the scaled fluxes to calculate a new spectro-magnitude. An example spectra before and after this scaling is shown in Fig. 5.3. The blue dots in the figure show the photometric fluxes computed from the Petrosian magnitudes to which the spectra were scaled. It can be seen that the original spectrum has a significantly different shape to what is expected given the magnitudes, and the scaled spectrum receives a tilt to correct this. There was a significant degree of variation in the magnitude of the scaling required for each spectrum. However, the offsets in the spectro- and photometric-scales appear to be mostly stochastic as the differences



(a) Original Spectrum



(b) Scaled Spectrum

Figure 5.3 An example spectra (a) before and (b) after the scaling described in the text. Black lines show the observed spectra and red the best-fit spectra recovered from VESPA. The blue dots show photometric fluxes computed from aperture matched Petrosian magnitudes. Black lines at the bottom of the panels indicate the SDSS filter transmission functions.

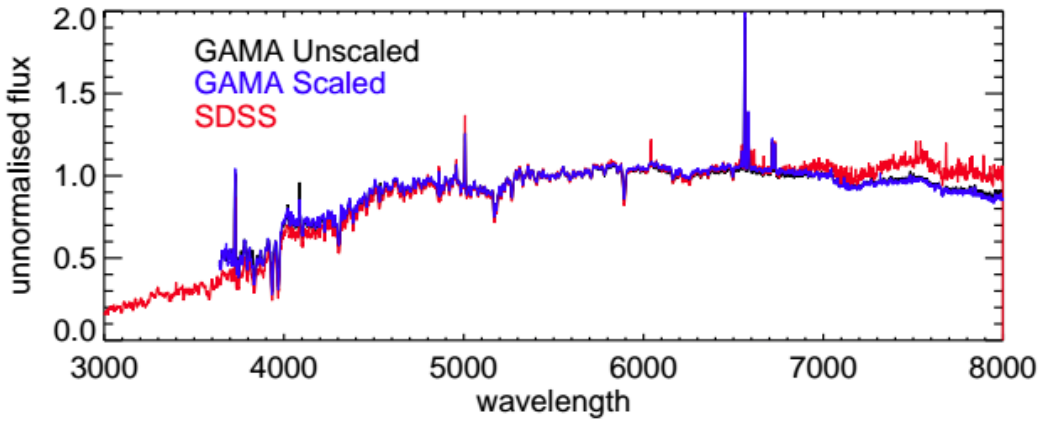


Figure 5.4 A stack of 5121 objects in SDSS and GAMA. In black, the original GAMA spectra, in blue, the scaled GAMA spectra and in red, the SDSS spectra. All spectra are normalised to unity at $\lambda = 5500$Å.

over a large set of spectra average out. Fig. 5.4 compares the stacked spectra of $\sim$ 5000 objects observed by both GAMA and SDSS to look for any systematic differences in shape. The black line shows the result of stacking the original GAMA spectra and the blue line the scaled GAMA spectra. It can be seen that the scaled and unscaled spectra are very similar when averaging over a large number of objects. The red line in the figure shows the stacked SDSS spectra, which displays a slightly different continuum shape. This possibly reflects radial colour gradients, with galaxies having redder nuclei on average. The lack of an average change of this sort when scaling GAMA spectra may seem to indicate that the effect is small, especially as 2dF fibres are smaller (2 arcsec diameter as against 3 arcsec for SDSS I/II). However, the poorer typical AAT seeing is the dominant effect: GAMA spectra will tend to be more representative of the mean spectrum as the result.

5.2 Quantifying star-formation histories

The GAMA spectra for all galaxies in the sample discussed in Section 3.1.1, after the aforementioned scaling to the photometry, were run through VESPA to recover the star formation histories (SFHs). The observed and best-fit spectra of two different galaxies, along with the recovered star formation histories are shown in Fig. 5.5 as an example. The bottom panels show the fraction of mass

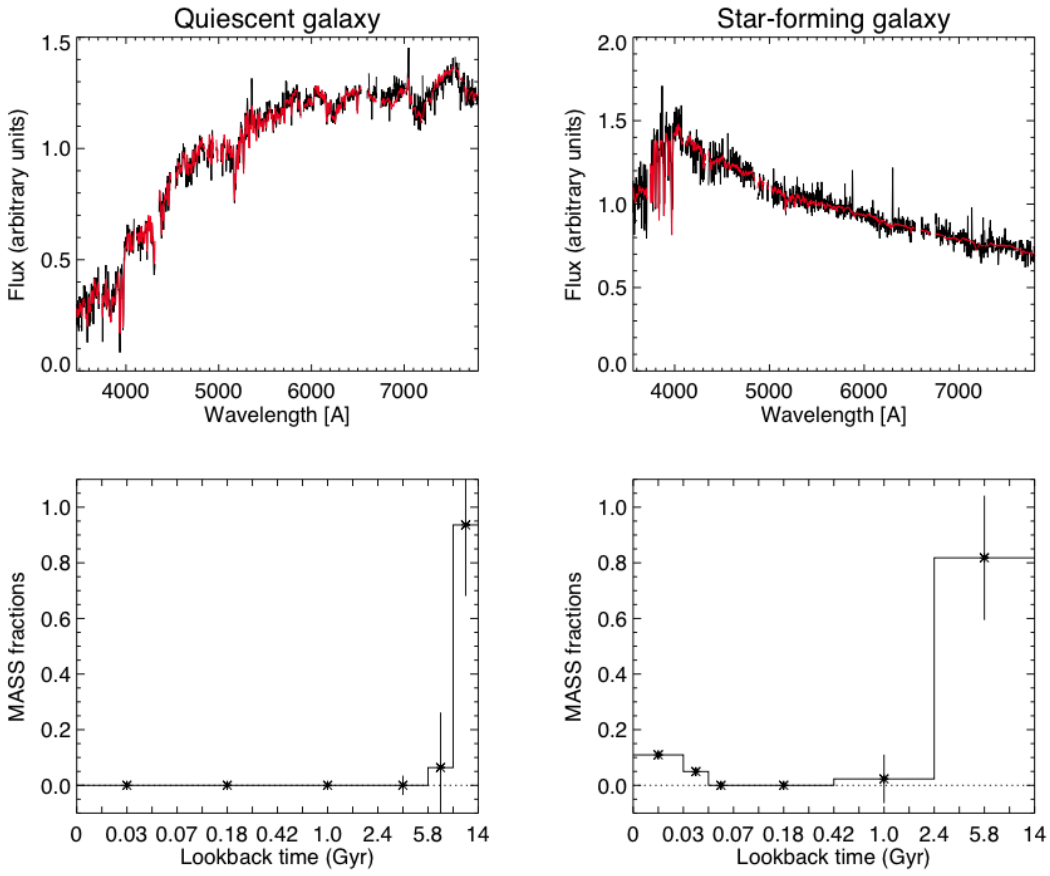


Figure 5.5 Example star formation histories recovered from VESPA. Top row shows spectra for two different galaxies: observed spectrum in black and the best-fit spectrum recovered from VESPA in red. The panels directly below the spectra show the recovered star formation histories.

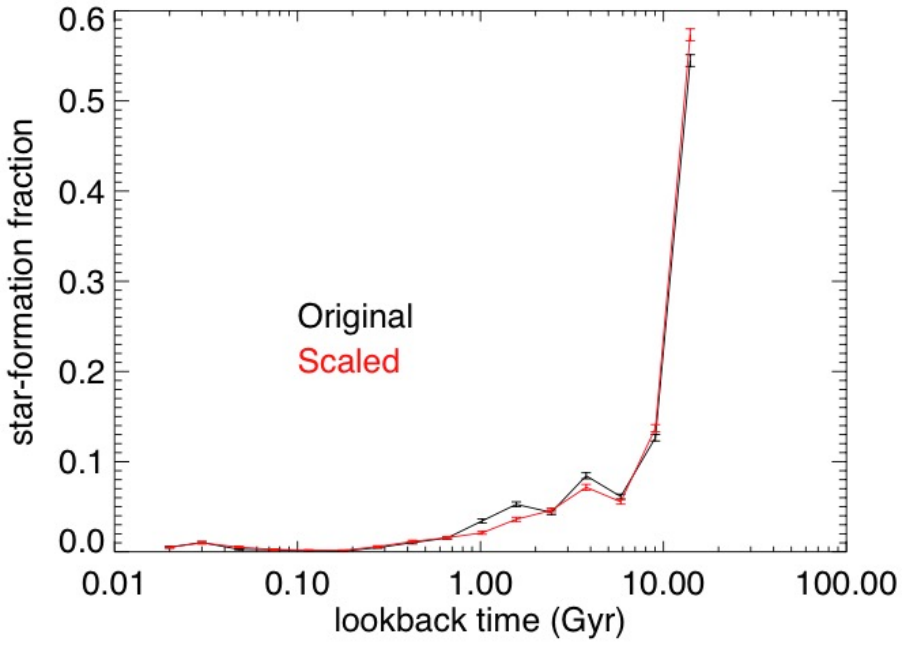


Figure 5.6 Combined star formation fraction as a function of lookback time for **black:** unscaled and **red:** scaled GAMA spectra.

formed within each time bin as recovered by VESPA. The star formation history of a given spectrum can only be recovered with a modest level of confidence, given the noise and uncertainties in data and limitations of the modelling. For that reason, the results of a large number of spectra can be 'stacked' to reduce uncertainties by combining a large number of results. For example, all objects within a given geometric environment can be stacked to produce the averaged star formation histories across the environment.

To investigate the effects of the scaling applied to the spectra on the recovered SFHs, both the scaled and original spectra were run through VESPA and the resulting SFHs stacked and compared, as shown in Fig. 5.6 (only the results were stacked, spectra were run through VESPA individually). It can be seen that although the scaling makes very limited difference to the stacked spectra, the stacked SFHs show a noticeable difference. It is a positive sign that the SFHs of the scaled spectra show a smoother curve than the unscaled spectra. The star formation histories of the scaled spectra are shown again in Fig. 5.7, this time with lookback time on a logarithmic scale, and with the spectra split by the mass of their hosting group, if any. The group catalogue introduced in Section

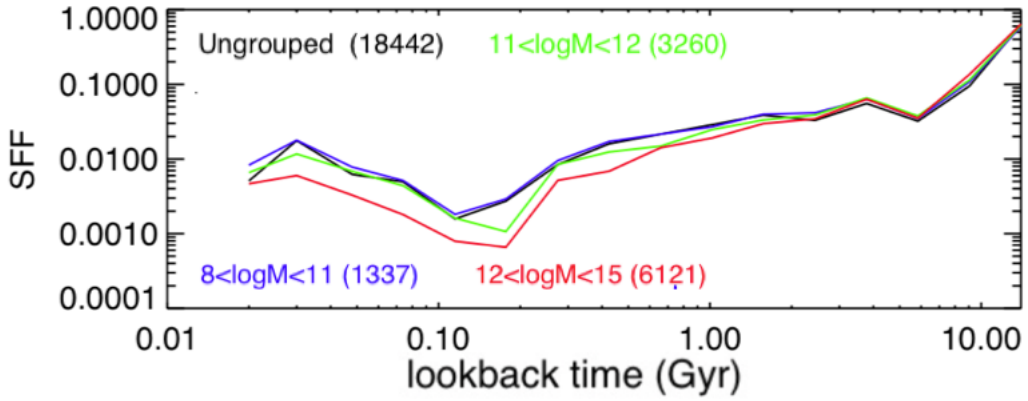
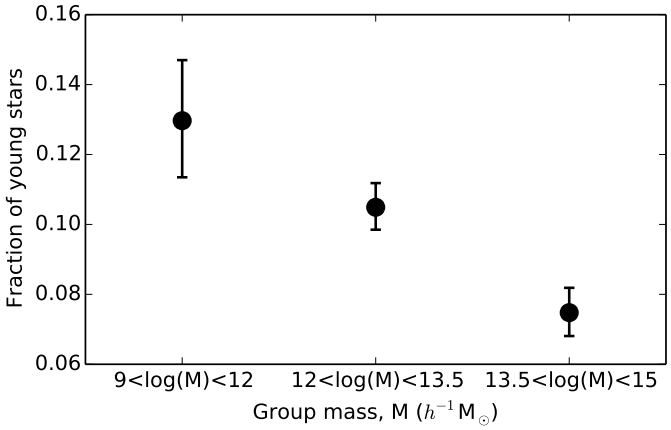


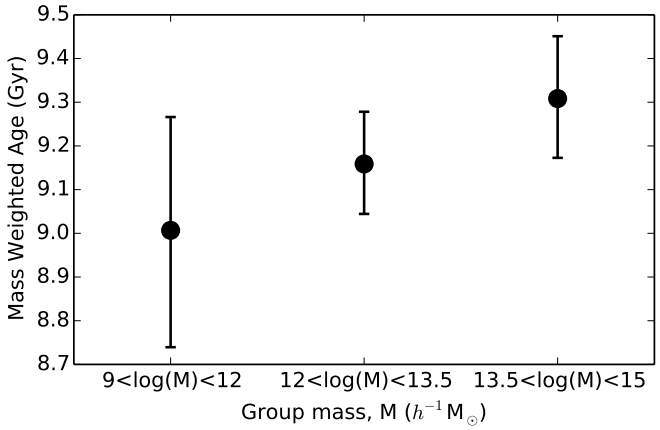
Figure 5.7 Average star formation fraction as a function of lookback time recovered from the scaled GAMA spectra split by the mass of their hosting group, if any. Coloured lines show the stacked star formation fractions of ungrouped galaxies **(black)**, and of galaxies in groups with group mass, M_{gr} ($h^{-1}M_{\odot}$), satisfying $8 < \log_{10}(M_{gr}) < 11$ **(blue)**, $11 < \log_{10}(M_{gr}) < 12$ **(green)** or $12 < \log_{10}(M_{gr}) < 15$ **(red)**. The number of galaxies within each category are shown by the coloured numbers in the key.

3.1.3 and the dynamical group mass estimates discussed in Section 4.2.1 are used for this categorisation. It can be seen that the recent star formation fraction is highest for the ungrouped galaxies, and then decreases with increasing group mass. The use of the logarithmic scale for the x-axis emphasises the upturn in the star formation for times ~ 0.1 Gyr or earlier, which is somewhat unexpected. It was found that this increase in star formation at recent times was affected by the inclusion of the u-band photometry, and it was not possible to include any more photometric information in order to test for further differences when more information was included. Therefore the detailed reliability of the star formation fractions recovered at recent times is questionable. For this reason, instead of analysing the full histories, there is a need to find one or more statistics that can robustly characterise the distinction between different star-formation histories.

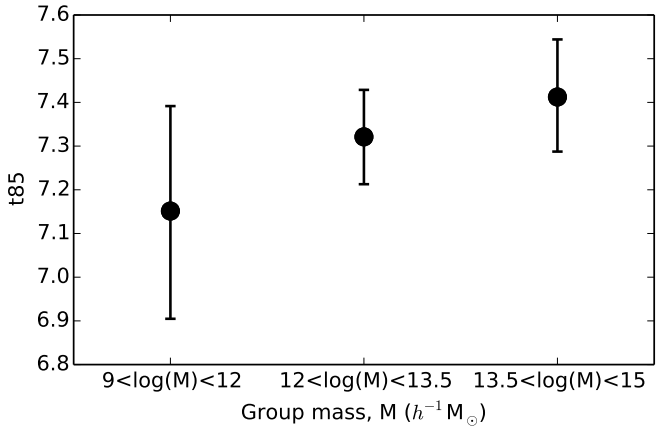
Three possible statistics were tested; firstly, the fraction of young stars in the galaxy, where young refers to stellar populations less than 275 Myr in lookback time in the galaxy's frame. Secondly, the mass weighted age of the galaxy and finally, t_{85}, the lookback time at which 85% of the mass of the galaxy had been formed. Since the later two statistics have dimensions of time, there is the question of relative to which frame the lookback time should be quoted. If the galaxy rest frame is used, a lookback time t for two galaxies observed at different redshifts will represent a different age of the Universe in our $z = 0$ observer's



(a) Young fraction



(b) Mass weighted age



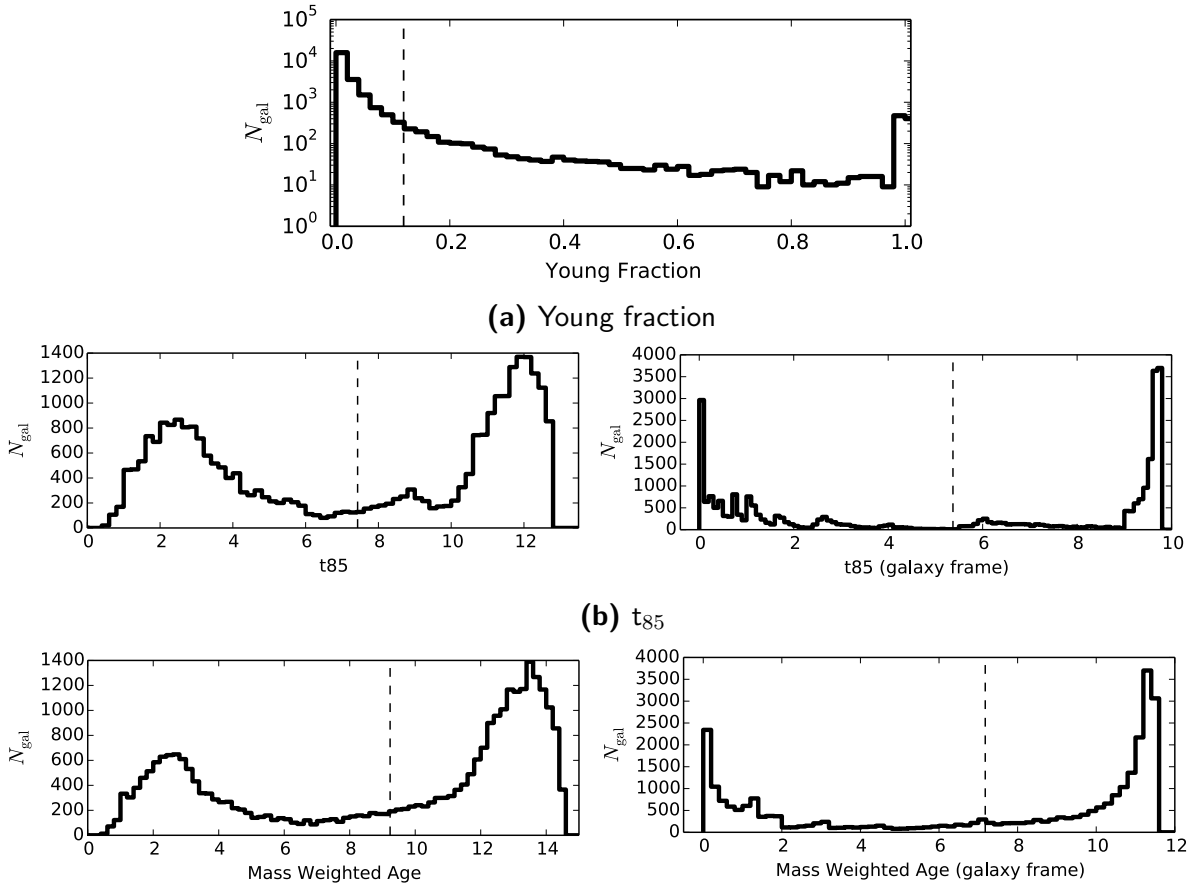
(c) t_{85}

Figure 5.8 Average value of the three statistics, **(a)** fraction of mass formed less than 275 Myrs ago, **(b)** the mass weighted age of the galaxy and **(c)** t_{85}, the lookback time at which 85% of the mass of the galaxy has been assembled, for galaxies split by the mass of their hosting galaxy, M (h^{-1}M$_\odot$). Error bars show bootstrapped 95% confidence intervals.

frame. It is possible to calculate the lookback time to the redshift of the observed galaxy and add this to the galaxy frame lookback time to convert the results to the common observer's frame. There are arguments in favour of both approaches depending on what is desired to be measured, therefore, where necessary, results from both frames are shown. Hereafter the statistics have been converted to the common observer's frame by the addition of the lookback time to the galaxy unless it is stated that the results are in the galaxy frame. The value of 85% was chosen by optimising the recovered formation times, t_x, on mock galaxies. A constant star formation rate was assumed across each individual bin in order for the mass fraction to be linearly interpolated across bins.

The average value of these 3 statistics was measured for the galaxies again split by the mass of their hosting group, as shown in Fig. 5.8, with error bars representing bootstrapped 95% confidence intervals. It can be seen that all statistics exhibit a roughly smooth, linear trend with group mass. The young fraction is seen to decrease with increasing group mass, indicating that more recent star formation has occurred in galaxies hosted by low mass groups. Consistently, the mass weighted age and t_{85} are seen to increase with increasing group mass, indicating that galaxies in higher mass groups are older. However the later two statistics are also consistent with no dependence on group mass, given the uncertainties.

As this is an ongoing project, the robustness of these statistics are still being investigated. Although the trends with mass are qualitatively as one might expect, their distributions raise questions as to the quality of the average value of the distribution as an indicator of formation time. Fig. 5.9 plots histograms of the resulting three statistics for all galaxies in the sample. The dashed vertical lines in the figure represent the average value of each statistic. The figure also shows the histograms of t_{85} and the mass weighted age in the galaxy rest frame. A strong bimodality can be seen in all statistics, and particularly strongly for t_{85}. Fig. 5.9a, the distribution of the fraction of young stars, shows a strong spike in the histogram for the last bin representing the significant proportion of galaxies with all star formation found to have occurred within the last 275 Myrs. These galaxies will contribute to the first bins of the t_{85} and mass weighted age histograms. Fig. 5.9b indicates that VESPA recovers very little star formation between 2 - 9 Gyrs in the galaxy rest frame. The smoother distributions seen



(a) Young fraction

(b) t_{85}

(c) Mass weighted age

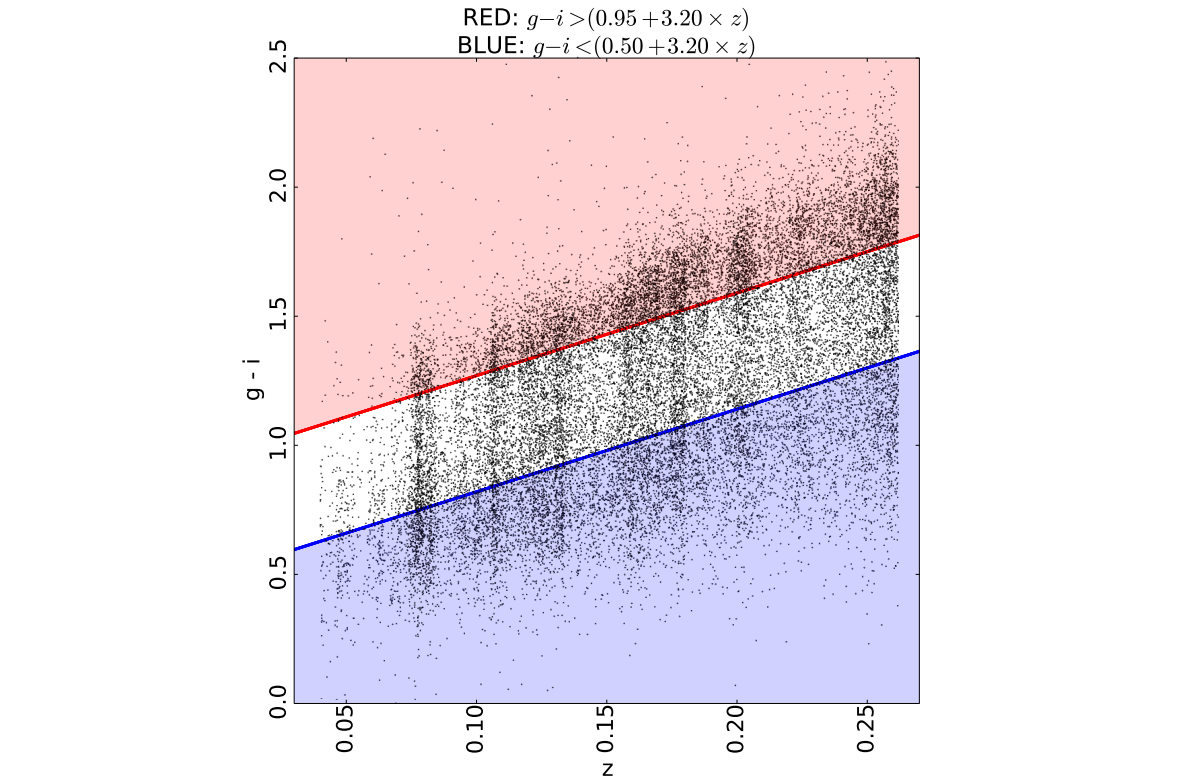
Figure 5.9 Distributions of the three statistics, **(a)** fraction of mass formed less than 275 Myrs ago, **(b)** t_{85}, the lookback time at which 85% of the mass of the galaxy had been assembled, **(c)** the mass weighted age of the galaxy, for all galaxies in the sample. For t_{85} and mass weighted age the left and right panels show the results in the observer's frame and the galaxy rest frame respectively. Dashed vertical lines show the average value of each statistic.

in the observers frame for these statistics are due to the redshift distribution of galaxies in the sample.

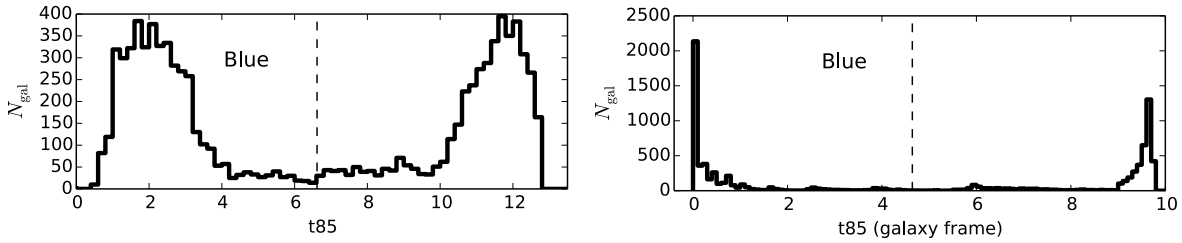
The task remains to identify what is causing this bimodality in the results. To investigate the bimodality further, with a focus on the t_{85} statistic, subsamples of galaxies were selected by colour, the signal to noise of the spectra, and amount of dust in the galaxy, as estimated by VESPA. Two subsamples of the galaxy population were selected from the extreme ends of the distribution for each of these three properties. Fig. 5.10 plots the distribution of t_{85} for red and blue galaxies selected by their $g-i$ colour. The $g-i$ colour and redshift of the galaxies in the sample are shown in Fig. 5.10a: the red and blue shaded regions indicate those galaxies selected for the red and blue samples respectively. It can be seen from Fig. 5.10b and Fig. 5.10c that the significant spike for $t_{85} \sim 0$ seen in the blue sample is not present in the red sample, which instead sees an increase in the number of high t_{85} galaxies.

Fig. 5.11 plots the distribution of t_{85} for galaxies selected according to the estimated abundance of dust. Fig. 5.11a shows a histogram of the recovered dust abundance for all galaxies in the sample, the pink and grey shaded regions indicate those galaxies selected for the low- and high-dust samples respectively. It can be seen from Fig. 5.11b and Fig. 5.11c that the spike for $t_{85} \sim 0$ seen in the dusty sample is significantly lower for the the low-dust galaxies, which instead sees an increase in the number of high t_{85} galaxies.

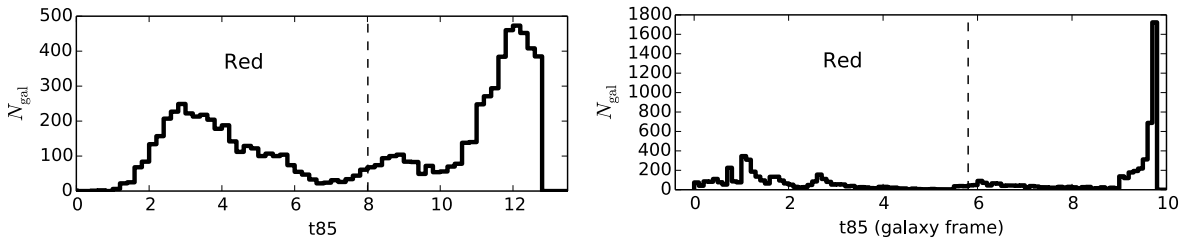
Fig. 5.12 plots the distribution of t_{85} for galaxies selected according to their signal to noise ratio (SNR). The distribution of SNR values for all galaxies in the sample is show in Fig. 5.12a, the red and green shaded regions indicate those galaxies selected for the low- and high-SNR samples respectively. It can be seen from Fig. 5.12b and Fig. 5.12c that the bimodality is significantly reduced for galaxies with higher signal to noise. Although less marked than the correlation with colour, this effect does indicate that the bimodality may, at least in part, be due to the inability of VESPA to recover accurate SFHs for the poor quality spectra. The SNR is also correlated with other properties of the galaxy. Fig. 5.13 shows how the fraction of galaxies falling into the colour and dust categories depicted by Fig. 5.10a and Fig. 5.11a varies when various SNR cuts are applied. It can be seen that higher SNR cuts select redder galaxies with less dust, characteristics which have been shown to reduce the bimodality



(a) g-i colour

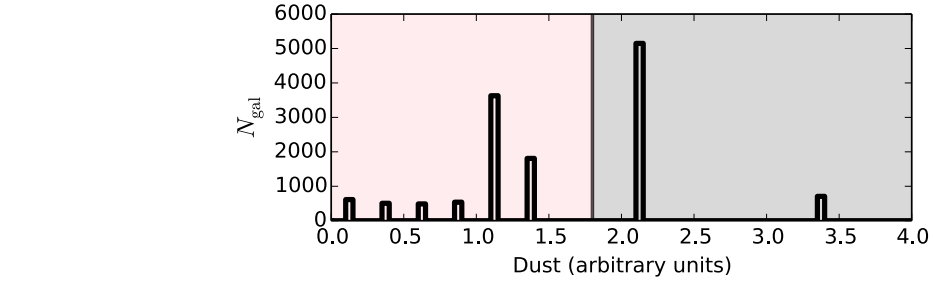


(b) t_{85}, blue galaxies

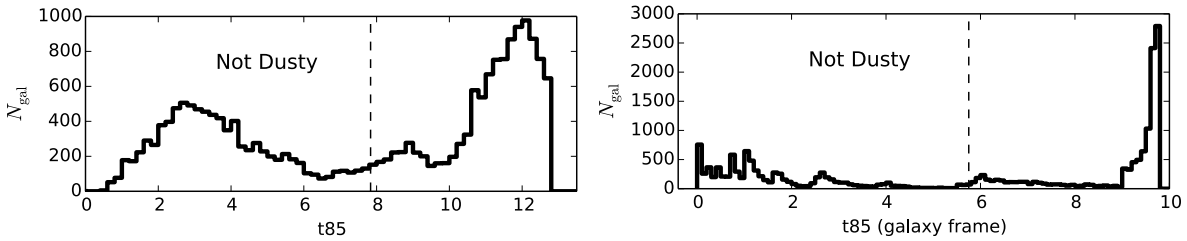


(c) t_{85}, red galaxies

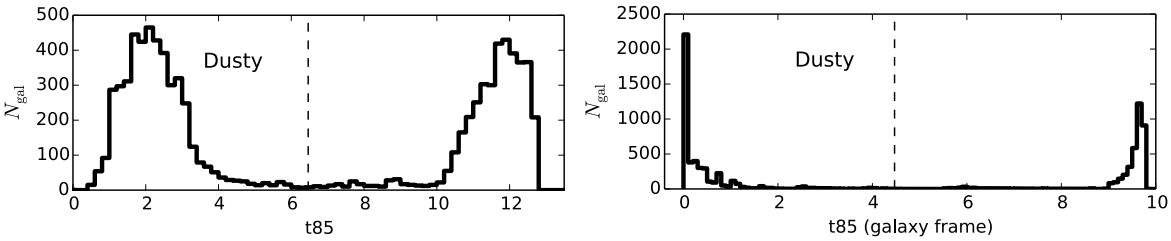
Figure 5.10 Distributions of the t_{85} statistic for red and blue galaxies. **(a)** g-i colour of galaxies in the sample as a function of redshift, illustrating the red and blue cuts applied to the data, as given above the panel. The blue sample is comprised of all galaxies below the blue line, and the red sample is comprised of all galaxies above the red line. **(b)** and **(c)**, distribution of t_{85} for the blue and red sample of galaxies, left and right panels show the results in the observer's frame and the galaxy rest frame respectively. Dashed vertical lines show the average value of each statistic.



(a) Abundance of dust for all galaxies in the sample

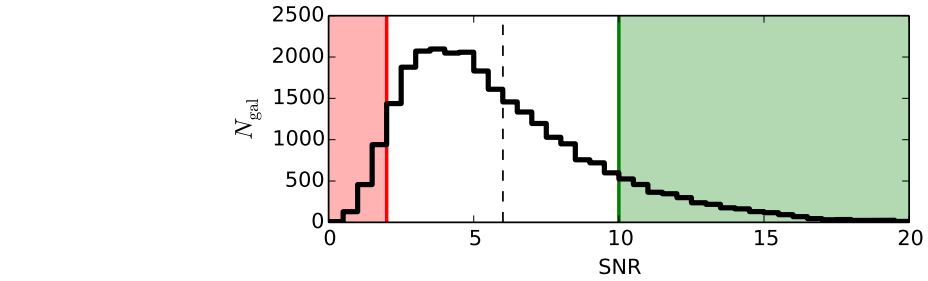


(b) t_{85}, non-dusty galaxies

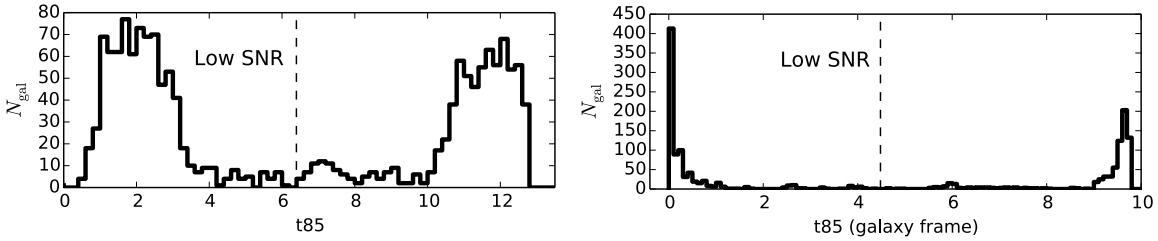


(c) t_{85}, dusty galaxies

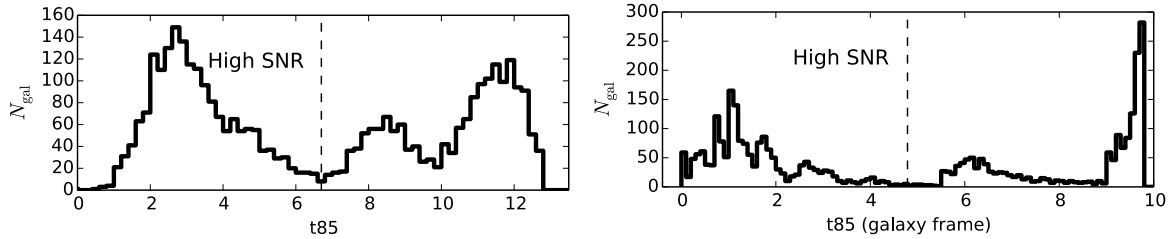
Figure 5.11 Distributions of the t_{85} statistic for galaxies with low and high dust abundance. **(a)** Histogram of the recovered dust for all galaxies in the sample. The low- and high-dust samples are comprised of all galaxies within the pink and grey shaded regions respectively. **(b)** and **(c)**, distribution of t_{85} for the low- and high-dust galaxies, left and right panels show the results in the observer's frame and the galaxy rest frame respectively. Dashed vertical lines show the average value of each statistic.



(a) signal-to-noise of all galaxies in the sample

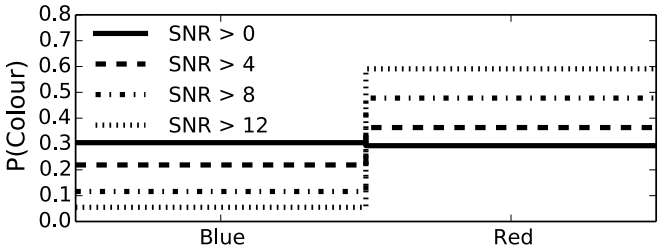


(b) t_{85}, Low signal-to-noise galaxies

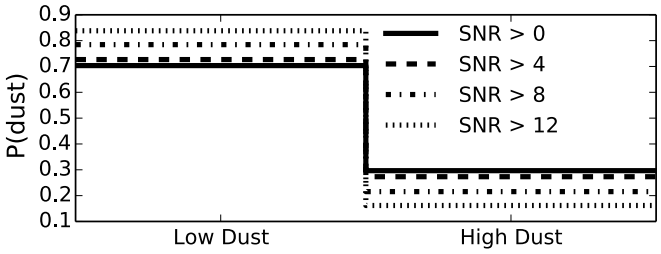


(c) t_{85}, High signal-to-noise galaxies

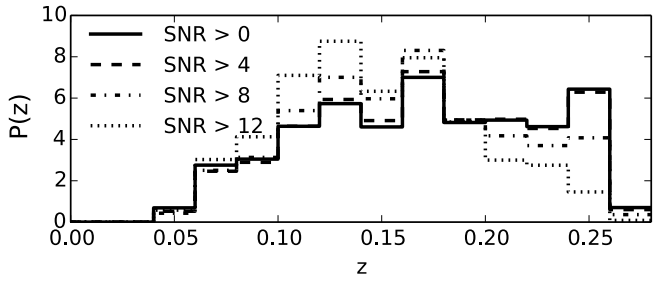
Figure 5.12 Distributions of the t_{85} statistic for galaxies with low and high signal to noise ratios (SNRs). **(a)** The distribution of SNR of all galaxies in the sample. The low and high SNR samples are comprised of galaxies within the red and green shaded regions respectively. **(b)** and **(c)**, distribution of t_{85} for the low and high SNR sample of galaxies, left and right panels show the results in the observer's frame and the galaxy rest frame respectively. Dashed vertical lines show the average value of each statistic.



(a) Colour



(b) Dust



(c) Redshift

Figure 5.13 Colour, dust and redshift of SNR selected subsamples. For galaxy subsamples selected according to the minimum SNR limits given in the key, the fraction of the subsample falling into the (**a**) colour and (**b**) dust categories of Fig. 5.10 and Fig. 5.11 respectively. (**c**) The redshift distribution of the SNR selected subsamples.

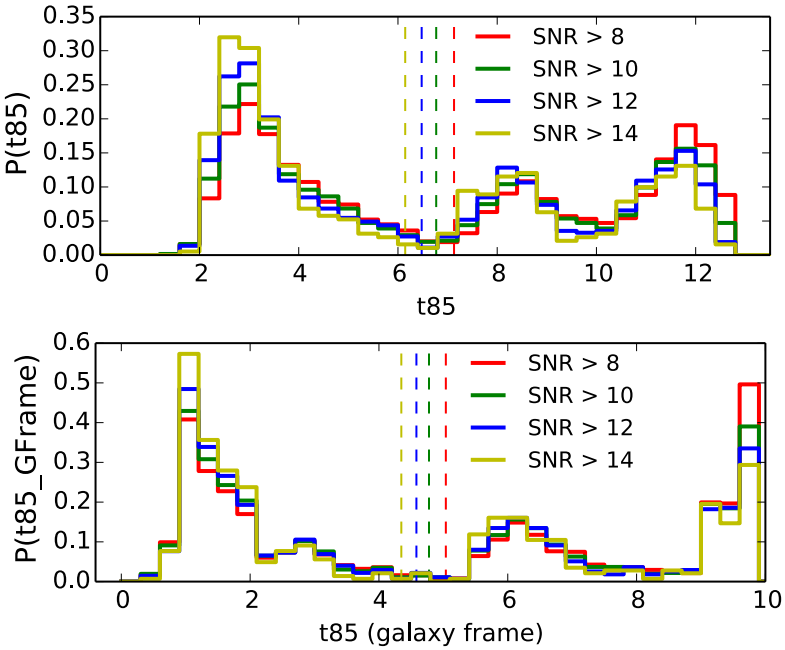


Figure 5.14 t_{85} distribution of red, low-dust galaxies with minimum SNR limits as given in the key. Top and bottom panels show the results in the observer's frame and the galaxy rest frame respectively. Dashed vertical lines show the average value of t_{85} for each subsample.

in the t_{85} distribution. The figure also shows the redshift distribution of galaxies with different minimum SNRs, and shows that higher SNR subsamples contain lower redshift galaxies. It may be hoped that, when including only red, low-dust galaxies, the results of high SNR cuts would show a convergence. However Fig. 5.14 shows the distributions of t_{85} for red, low-dust galaxies with various minimum SNR limits and no such convergence is seen. The dashed line in the figure shows the average value of the distribution and it can be seen that, on average, a lower value of t_{85} is recovered for higher SNR galaxies even when only red, low-dust galaxies are considered.

5.3 Star-formation histories within the observed cosmic web

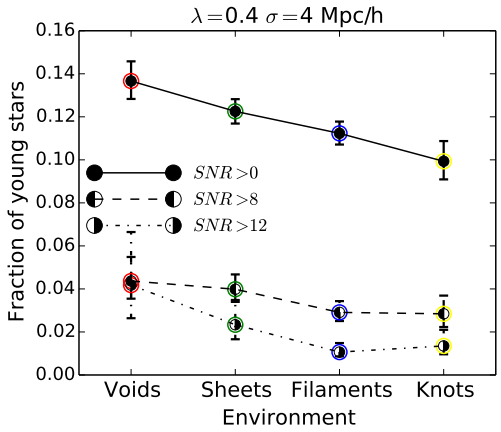
Notwithstanding the discussion of robustness in the previous section, this section presents an initial analysis of the variation of stellar-formation times within the

cosmic web. Using the geometric environment classifications presented in Section 3.3, the average value of the three statistics introduced in Section 5.2 for the four components of the cosmic web can be calculated and compared. These are shown in Fig. 5.15 and Fig. 5.16 for geometric environments defined with the parameter sets $(\sigma, \lambda_{\rm th}) = (4\,h^{-1}{\rm Mpc}, 0.4)$ and $(10\,h^{-1}{\rm Mpc}, 0.1)$ respectively. The average value of all three statistics is seen to decrease with environment from voids to knots (although this trend is weak for the fraction of young stars within environments defined with the larger smoothing scale). This, somewhat paradoxically, implies that high dimensional environments (e.g. voids and sheets) are simultaneously both older and younger than lower dimensional environments (knots and filaments). A higher young star fraction implies a younger age, due to the larger young population within the galaxy, whilst a higher t_{85} or mass weighted age implies an older age.

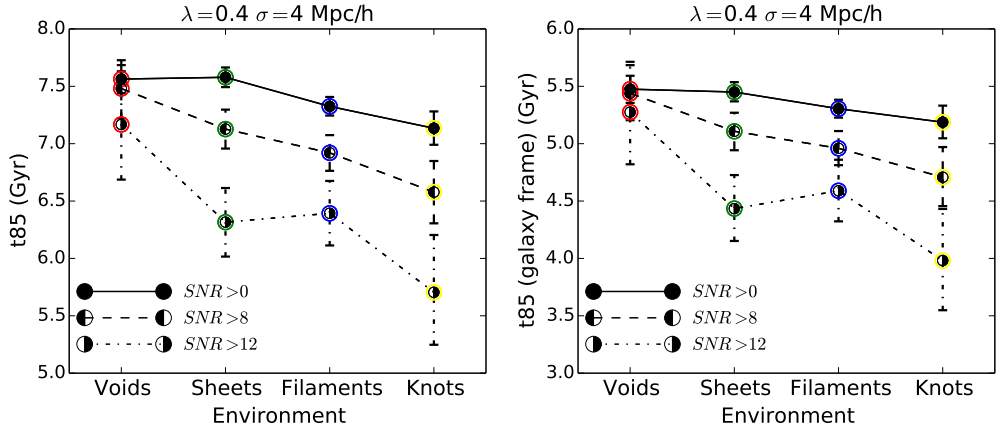
The figures also show the results for the subsets of galaxies with SNRs greater than 8 and 12. This SNR cut introduces the largest shift for the average fraction of young stars in the galaxies. This is due to the removal of a significant portion of low SNR galaxies with all of their recovered star formation found to have occurred within the past 275 Myrs. It can be seen that the trends with environment persist, and arguably become stronger, when low-SNR galaxies are removed

The distributions of the three statistics from which these average values are derived are shown in Figs. 5.17-5.19. Only results for the geometric environments derived with the smaller smoother scale parameter set are shown here, though it is noted that little difference is seen between these and the corresponding plots for the larger smoothing scale environments. Qualitatively the distributions are similar across all environments. When removing galaxies with low SNR, a larger difference between the distributions of the environments appears for the t_{85} and mass weighted age statistics, where a sharper peak around ~ 2 Gyr is seen for the lower dimensional environments.

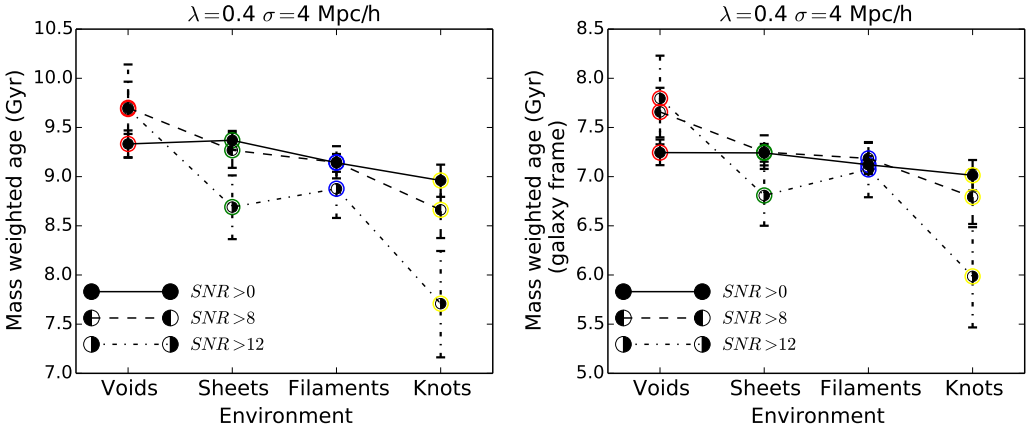
It should be noted that the opposite trends of age with environment implied by the fraction of young stars and by t_{85} or the mass weighted ages are not necessarily inconsistent with each other, due to the difference in the statistics. For example, a set of galaxies with most of their star formation having occurred at high lookback time, and the remainder having occurred very recently would display both a higher t_{85} and a higher fraction of young stars than a set of galaxies



(a) Fraction of young stars

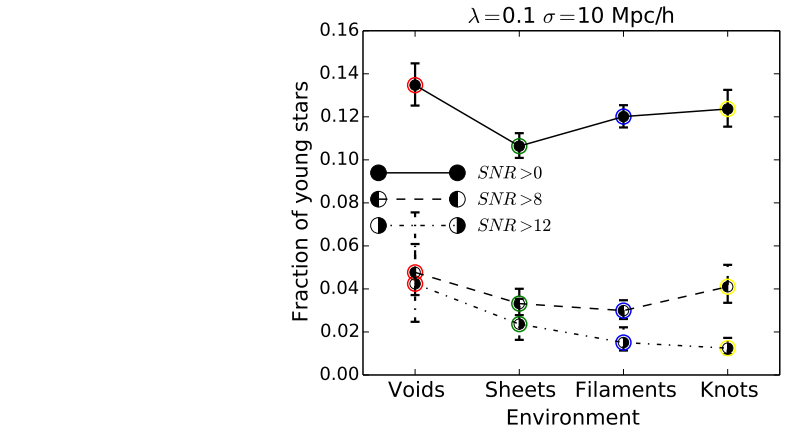


(b) t_{85}

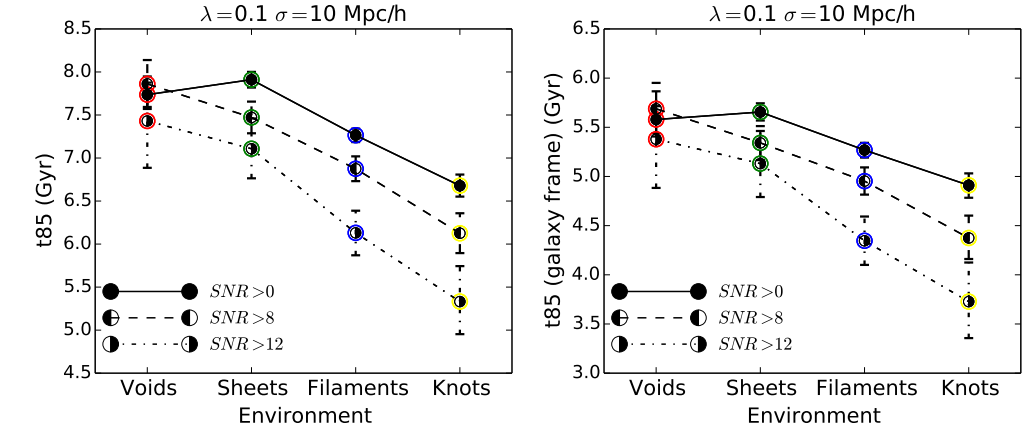


(c) Mass weighted age

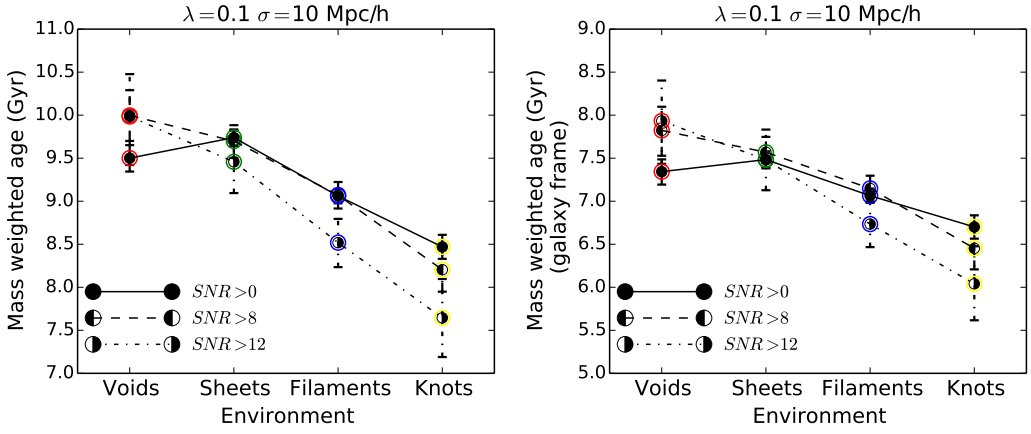
Figure 5.15 Average value of the three statistics, **(a)** fraction of young (< 275Myr) stars, **(b)** t_{85} and **(c)** mass weighted age within the four components of the cosmic web, defined with $(\sigma, \lambda_{\rm th}) = (4\,h^{-1}{\rm Mpc}, 0.4)$. For **(b)** and **(c)**, left and right panels show the results in the observer's frame and the galaxy rest frame respectively. Solid circles show results for the full galaxy sample whilst the half filled circles connected with dashed and dot-dashed lines show results for the subsample of galaxies with $SNR > 8$ and 12 respectively. Error bars show bootstrapped 95% confidence intervals.



(a) Fraction of young stars



(b) t_{85}



(c) Mass weighted age

Figure 5.16 Average value of the three statistics, **(a)** fraction of young (< 275Myr) stars, **(b)** t_{85} and **(c)** mass weighted age within the four components of the cosmic web, defined with $(\sigma, \lambda_{\rm th}) = (10\,h^{-1}{\rm Mpc}, 0.1)$. For **(b)** and **(c)**, left and right panels show the results in the observer's frame and the galaxy rest frame respectively. Solid circles show results for the full galaxy sample whilst the half filled circles connected with dashed and dot-dashed lines show results for the subsample of galaxies with $SNR > 8$ and 12 respectively. Error bars show bootstrapped 95% confidence intervals.

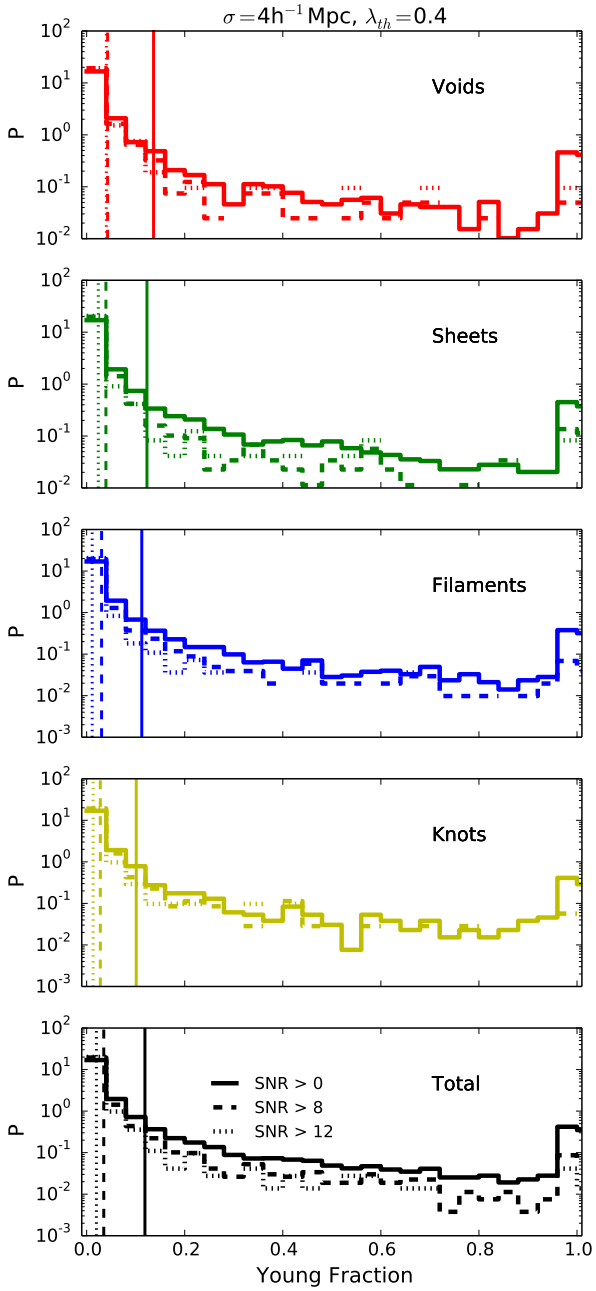


Figure 5.17 The distribution of the fraction of young (< 275 Myr) stars within galaxies selected from the four components of the cosmic web, defined with $(\sigma, \lambda_{\rm th}) = (4\,h^{-1}{\rm Mpc}, 0.4)$, with the combined distribution of all environments in the bottom panel. Solid lines indicate the distribution of all galaxies, dashed- and dotted-lines those galaxies with an SNR > 8 and 12 respectively. The vertical lines in each panel indicate the average value of the distributions.

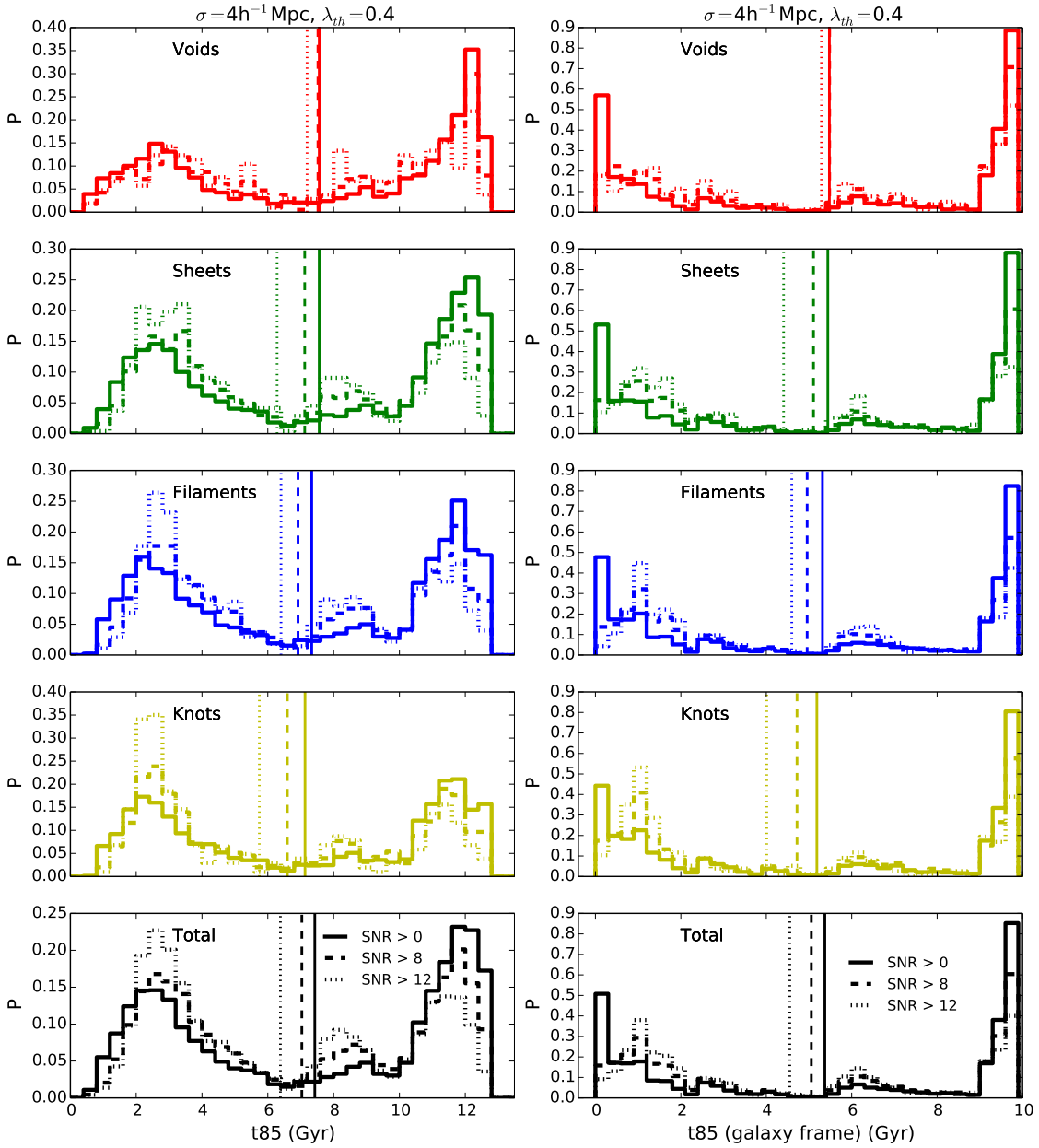


Figure 5.18 The distribution of t_{85} of galaxies selected from the four components of the cosmic web, defined with $(\sigma, \lambda_{\rm th}) = (4\,h^{-1}{\rm Mpc}, 0.4)$, with the combined distribution of all environments in the bottom panel. Solid lines indicate the distribution of all galaxies, dashed- and dotted-lines those galaxies with an SNR > 8 and 12 respectively. The vertical lines in each panel indicate the average value of the distributions. Left and right panels show results in the observer's and galaxy frame respectively.

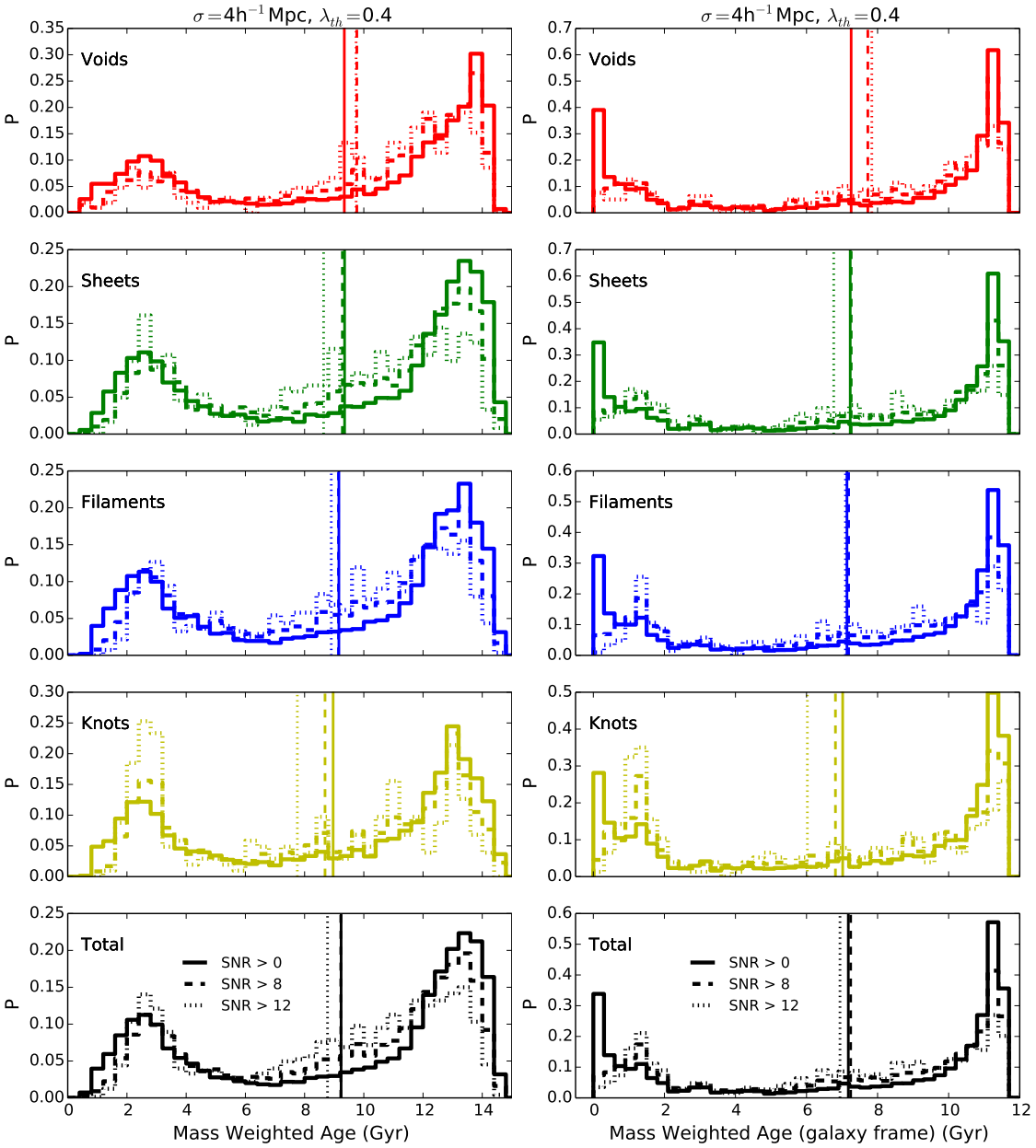


Figure 5.19 The distribution of the mass weighted ages of galaxies selected from the four components of the cosmic web, defined with $(\sigma, \lambda_{\rm th}) = (4\,h^{-1}{\rm Mpc}, 0.4)$, with the combined distribution of all environments in the bottom panel. Solid lines indicate the distribution of all galaxies, dashed- and dotted-lines those galaxies with an SNR > 8 and 12 respectively. The vertical lines in each panel indicate the average value of the distributions. Left and right panels show results in the observer's and galaxy frame respectively.

with all star formation having occurred at moderate lookback times.

One of the aims of this chapter is to look for evidence of assembly bias, which would manifest itself as a dependence of stellar-formation time on geometric environment for fixed group mass. For this reason, the variation in these statistics within different regions of the cosmic web is calculated for galaxies in groups of similar mass, $M_{\rm group}$ (here the dynamical mass is used). Figs 5.20-5.25 show, for each statistic and geometric environment parameter set, the average value of the statistic for ungrouped galaxies, and for galaxies in groups of mass within a given range, within different regions of the cosmic web. Here error bars show bootstrapped 95% confidence intervals, however it should be noted that these errors are expected to be an underestimate of the total uncertainties. The grey bands in the figures indicate the average values and confidence intervals for galaxies from all geometric environments. The fraction of young galaxies appears to be consistent with no dependence on geometric environment, for fixed group mass. However, a dependence on geometric environment remains for t_{85} and mass weighted age, particularly for galaxies from groups with mass $13 < \log_{10}(\mathrm{M_{group}}) < 14$.

5.4 Discussion

This chapter has presented the initial results of an investigation into the variation of stellar-formation times within the cosmic web. VESPA, the full spectral-fitting code, was used to extract star formation histories from GAMA spectra scaled to match SDSS optical photometry. Given the limited robustness of the full histories, three statistics indicative of stellar-formation time were presented; t_{85}, mass weighted age, and the fraction of young stars. It was shown that each of these three statistics implied that galaxies from higher mass groups formed earlier. The results suggest that, on average, galaxies from higher mass groups have a smaller fraction of young stars and hence less recent star formation. Similarly, the average values of t_{85} and of the mass weighted age of galaxies were shown to increase for galaxies selected from higher mass groups. This indicates that the majority of star formation occurs earlier for galaxies from high mass groups than from low mass groups.

A deeper look at the values of the statistics for galaxies in the sample showed

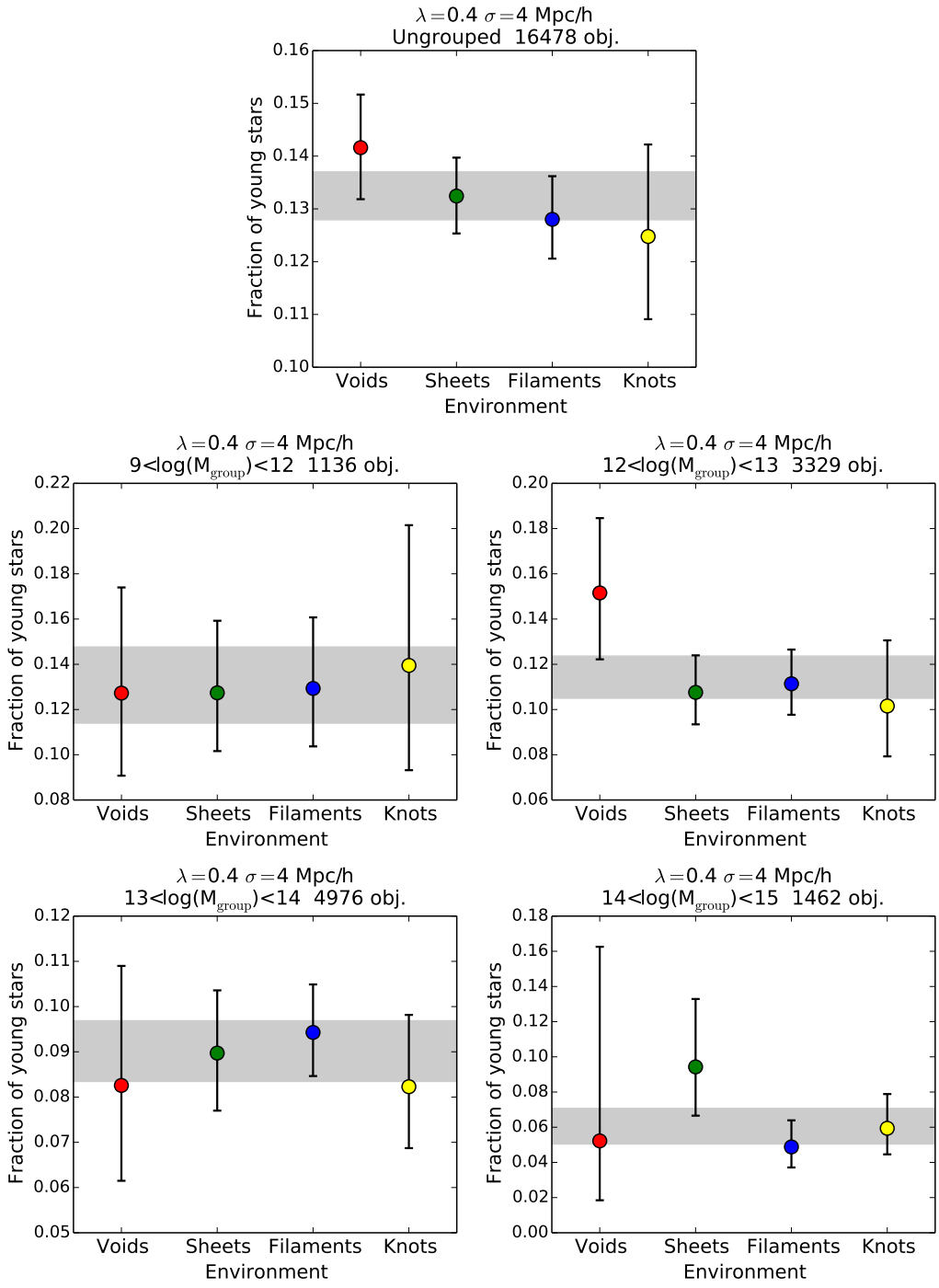


Figure 5.20 Average fraction of young (< 275 Myr) stars of galaxies from the four components of the cosmic web, defined with $(\sigma, \lambda_{\rm th}) = (4\,h^{-1}{\rm Mpc}, 0.4)$, for galaxies selected according to the dynamical mass of their hosting group, $\mathrm{M_{group}}$, if any, as stated above each panel. Error bars show bootstrapped 95% confidence intervals. Grey bands indicate the average value of galaxies from all geometric environments and its corresponding uncertainty.

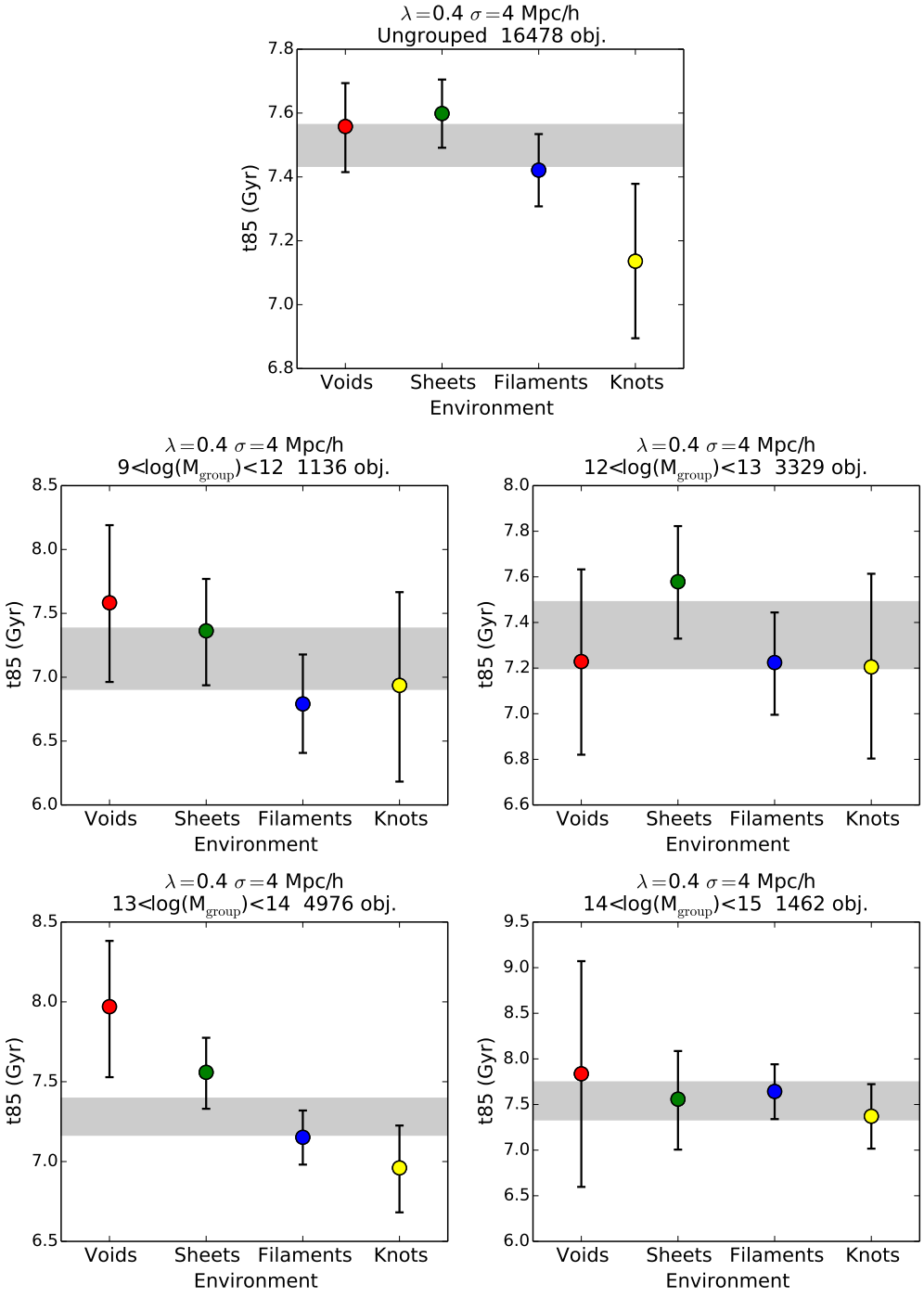


Figure 5.21 Average value of t_{85} within the four components of the cosmic web, defined with $(\sigma, \lambda_{\rm th}) = (4\,h^{-1}{\rm Mpc}, 0.4)$, for galaxies selected according to the dynamical mass of their hosting group, $\rm M_{group}$, if any, as stated above each panel. Error bars show bootstrapped 95% confidence intervals. Grey bands indicate the average value of galaxies from all geometric environments and its corresponding uncertainty.

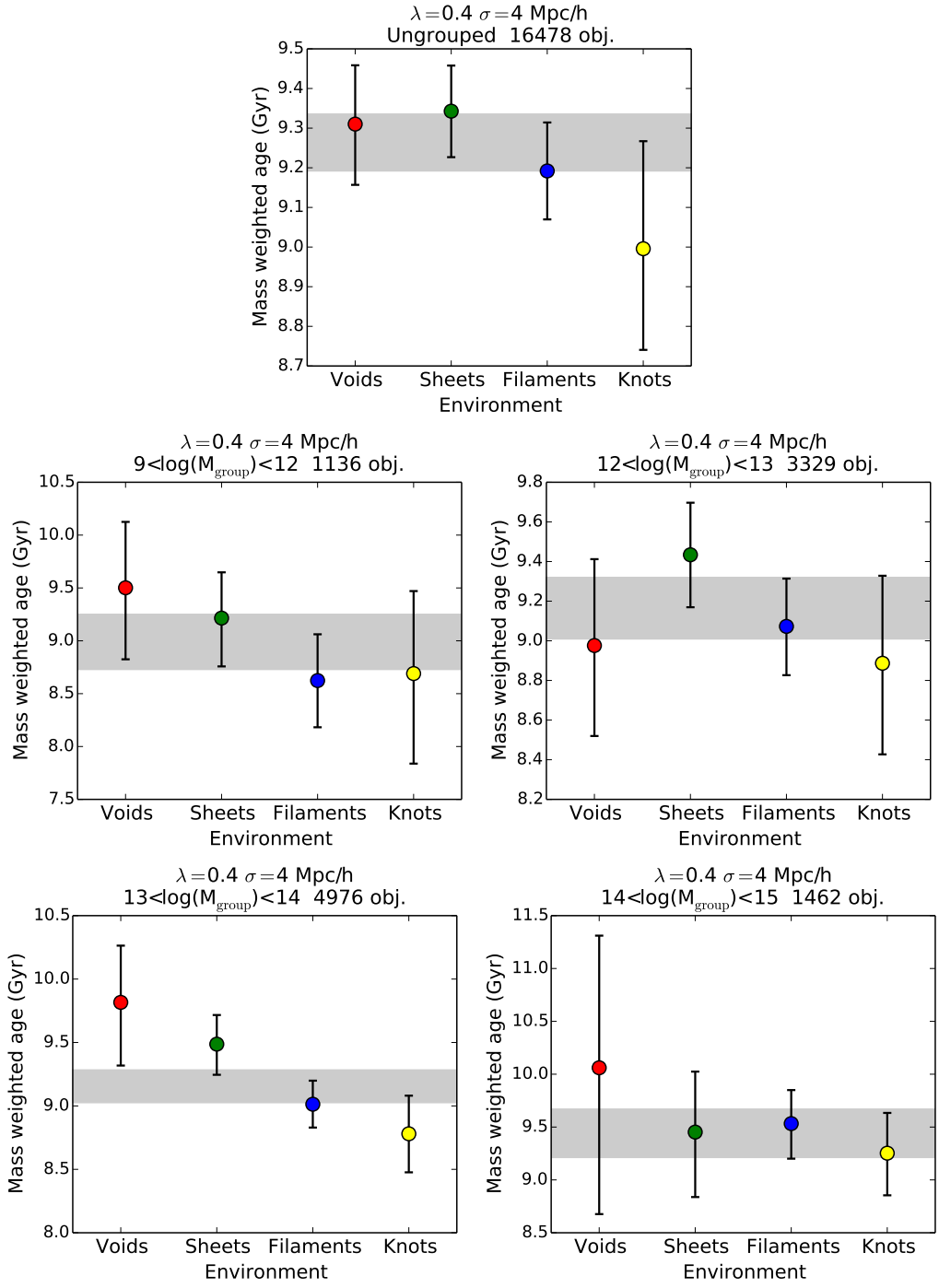


Figure 5.22 Average value of mass weighted age within the four components of the cosmic web, defined with $(\sigma, \lambda_{\rm th}) = (4\,h^{-1}{\rm Mpc}, 0.4)$, for galaxies selected according to the dynamical mass of their hosting group, $\mathrm{M_{group}}$, if any, as stated above each panel. Error bars show bootstrapped 95% confidence intervals. Grey bands indicate the average value of galaxies from all geometric environments and its corresponding uncertainty.

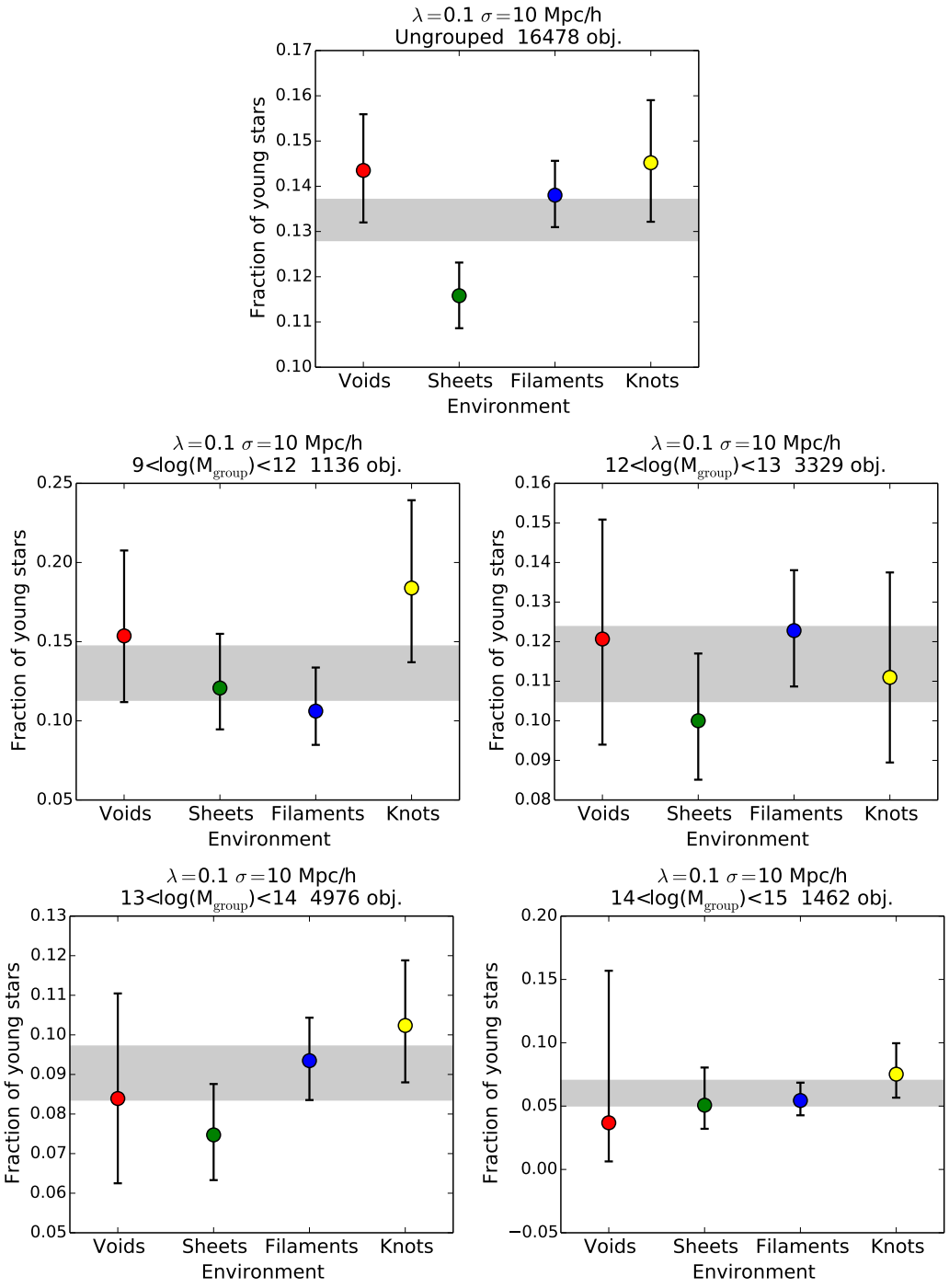


Figure 5.23 Average fraction of young (< 275 Myr) stars within galaxies from the four components of the cosmic web, defined with $(\sigma, \lambda_{\rm th}) = (10\,h^{-1}{\rm Mpc}, 0.1)$, for galaxies selected according to the dynamical mass of their hosting group, $\mathrm{M_{group}}$, if any, as stated above each panel. Error bars show bootstrapped 95% confidence intervals. Grey bands indicate the average value of galaxies from all geometric environments and its corresponding uncertainty.

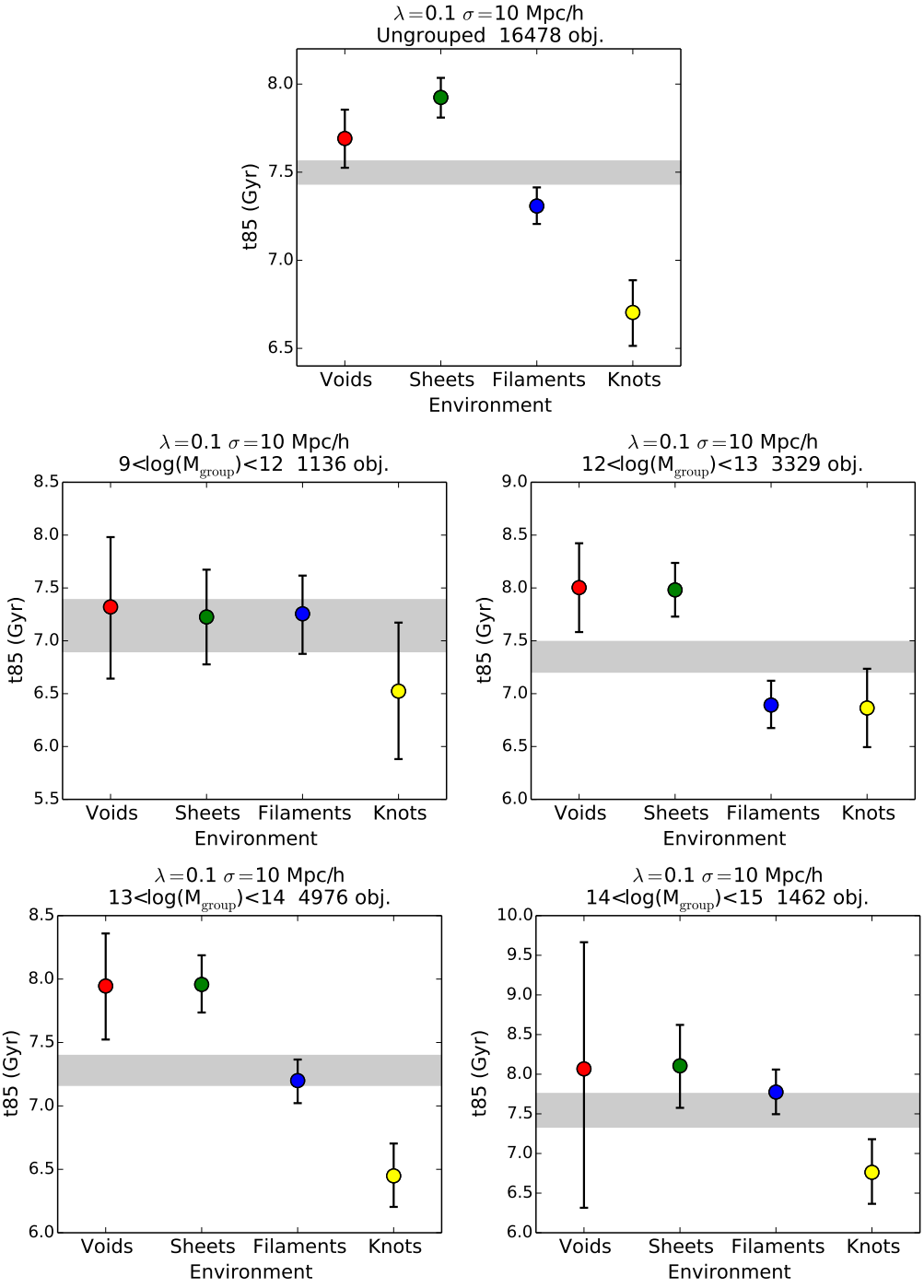


Figure 5.24 Average value of t_{85} within the four components of the cosmic web, defined with $(\sigma, \lambda_{\rm th}) = (10\,h^{-1}{\rm Mpc}, 0.1)$, for galaxies selected according to the dynamical mass of their hosting group, $\rm M_{group}$, if any, as stated above each panel. Error bars show bootstrapped 95% confidence intervals. Grey bands indicate the average value of galaxies from all geometric environments and its corresponding uncertainty.

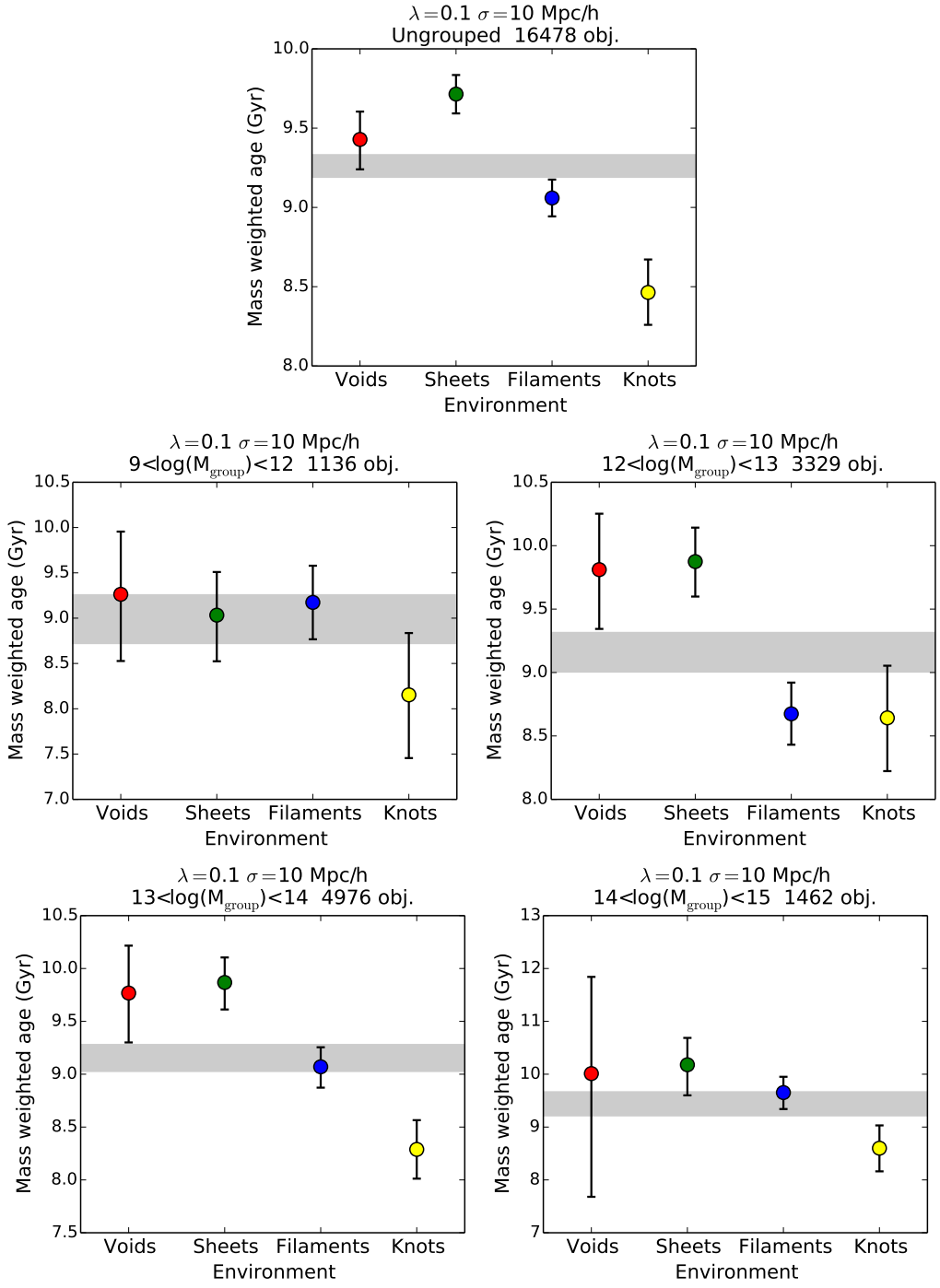


Figure 5.25 Average value of mass weighted age within the four components of the cosmic web, defined with $(\sigma, \lambda_{\rm th}) = (10\,h^{-1}{\rm Mpc}, 0.1)$, for galaxies selected according to the dynamical mass of their hosting group, $\mathrm{M_{group}}$, if any, as stated above each panel. Error bars show bootstrapped 95% confidence intervals. Grey bands indicate the average value of galaxies from all geometric environments and its corresponding uncertainty.

bimodal distributions, with a significant fraction of the galaxies having estimated formation times at one of the extreme ends of the scale. It was found that this bimodality was a feature of the blue, dusty, or low SNR galaxy populations in particular. Without reliable confidence limits on the recovered star formation histories, it is difficult to ascertain whether this bimodality is a true physical feature of these galaxy populations or, rather, is due to an inability of VESPA to accurately recover their histories. However, even if issues such as the low SNRs have some impact on the detailed results, the bimodality does strongly correlate with galaxy colour.

Nonetheless, when the variation of these estimated stellar-formation times across the four components of the cosmic web was investigated, a strong trend was seen with all three statistics decreasing with geometric environment from voids to knots. For the fraction of young stars, a decrease from voids to knots suggests that void galaxies have, on average, larger young stellar populations and more recent star formation than the lower dimensional environments (the reader is reminded that 3D voids are the highest dimensional environment, and knots have the lowest dimensionality). However, for t_{85} and the mass weighted ages, there is an apparently contradictory decrease, suggesting that most of the star formation occurred at earlier times in voids than in knots. These two results need not necessarily be in conflict. From the point of view of the separate universe model, voids are thoroughly Λ-dominated. It therefore makes sense that the bulk of structure formation within them would have frozen out at high redshift, leading to a larger t_{85} and mass weighted age. But this early freeze-out shifts the halo mass function to low masses. This is presumably the explanation for the well-known fact that voids are richer in dwarf galaxies (Croton et al. 2005) – but such galaxies generally show active current star-formation, as manifested by the high young fraction. In short, the two measures refer to different phases of a galaxy's life.

To search for assembly bias it is necessary to disentangle the variations of stellar-formation time with halo mass from those with the geometric environments. The fraction of young stars in galaxies from groups of similar mass was seen to be consistent with no dependence on geometric environment and hence displayed no evidence of assembly bias. However, when looking at t_{85} and the mass weighted age of galaxies from groups of similar mass, some dependence on

geometric environment remains. The limited quantification of all uncertainties at present restricts the ability to test whether these results are still consistent with no assembly bias, though these preliminary results currently indicate that assembly bias may exist.

As this chapter details an ongoing project, a fuller analysis of the robustness of these results is left for future work. The effect of the stellar population models and of the dust models on the recovered star formation histories remains to be fully tested. Additionally, the variation in the redshift distribution of galaxies from different environments may need to be considered when comparing the stellar-formation times. It may also be beneficial to calculate statistics for all galaxies by considering only star-formation that occurred before the lookback time to the highest redshift galaxy in the sample – this would remove any issues arising from the fact that there is no information on the star-formation that has occurred since observation, and the time-scale of this unknown era varies according to the redshift of the galaxy. It is also possible that a more optimal subset of galaxies, for example only those galaxies whose spectra have a high SNR, exists. Another potential approach is to stack some number, n, of spectra from the same environment or group *before* running through VESPA. Due to the non-linear nature of the dust extinction modelling, stacking all of the spectra before computing the SFHs will not produce the same results as stacking SFHs recovered from the individual spectra. However there may be some optimal number of spectra whose stacked spectrum has a high enough SNR to produce more robust results. This remaining work notwithstanding, this chapter attempts to address an interesting issue in a novel way, and has the potential to extract valuable information on the influence of large-scale environments on galaxy formation.

Chapter 6

Conclusions

This thesis has investigated the impact of geometric environments, as defined by the components of the cosmic web, on large-scale structure. This work was motivated by the need for a better understanding of the role of the environment during the formation and evolution of LSS. The aim was to search for any modulating effect of the cosmic web on galaxy and halo properties by utilising both galaxy redshift surveys and data from numerical simulations. The main conclusions, limitations and scope for further work within each investigation are now discussed below, followed by final concluding remarks.

6.1 The conditional halo mass function

Chapter 2 focused on simulated data and utilised the MDR1 dark matter only simulation. It was shown how the geometric environments of the full volume can be classified by application of the tidal tensor prescription. A practically motivated choice of the eigenvalue threshold, $\lambda_{\rm th}$, a free parameter of the model, was suggested whereby $\lambda_{\rm th}$ is set so as to obtain optimal statistics in each environment. It was shown that this choice can lead to classifications in good agreement with the visual impression of the cosmic web.

It would be interesting to investigate how the cosmic web defined through the tidal tensor prescription compares to the web as defined by other approaches. For example, by comparing the results presented here with those derived from the velocity shear prescription, which has been said to outperform the tidal tensor approach on small scales. Numerous other approaches have been proposed and

a better understanding of the extent of their agreement would be beneficial. However, the inapplicability of most of these methods to observed galaxy distributions meant that such a study was not a priority here. Some extensions to the tidal tensor prescription have employed a multi-scale approach, allowing a range of scales to be probed. Whether taking such an approach here has any influence on the main conclusions of this chapter would be interesting to investigate.

The aim of this chapter was to investigate the dependence of the halo mass function on location within the cosmic web, in order to search for any influence of the tidal forces associated with the web on halo properties and abundance. Current theoretical models assume no such dependencies beyond that of the large-scale mean density and this investigation was intended to test the validity of this assumption. Conditional halo multiplicity functions were calculated for the four components of the cosmic web. These were shown to display significantly different amplitudes, particularly for high masses.

In conjunction with this, Chapter 2 showed how Gaussian theory can make predictions on the halo mass function within different components of the cosmic web, and compared these theoretical results to those obtained empirically from the simulation. Two theoretical approaches were used: the excursion set formalism and the effective-universe approach. Neither of these approaches were able to yield quantitatively precise results though they were both in qualitative agreement with the simulated mass functions. The discrepancies were greatest for small smoothing scales and large $\lambda_{\rm th}$, when the density field is further from the linear regime. The disagreement can be understood from the fact that the standard prediction for the conditional mass function is only valid for low densities, i.e. below the collapse threshold, and for masses below the filter scale.

The theoretical mass functions are tightly linked to the density distributions of the environment. It was shown that the Gaussian theory poorly predicts the density distributions of the geometric environments measured directly from the simulation. Again, this can be attributed to the nonlinearity of the density field on the scales that were probed. Although de-lognormalising the density field showed some improvement it still did not produce good agreement. This limited the success of a fully theoretical approach and led to information of the simulated density distributions being required for the theoretical predictions. It

was noted that the Gaussian approach shows only an approximate understanding of the trends in halo properties with environment. Calibrating from numerical simulations may potentially lead to more accurate predictions.

The fundamental prediction of the Gaussian theory work was that different geometric environments with the same overdensity should have identical halo mass functions. That is, for a given overdensity, the halo mass function is independent of location within the cosmic web. This was verified in the MDR1 simulation to a good approximation for a wide range of masses, filters and eigenvalue thresholds. It was shown that all differences between the simulated halo mass functions of different web-environments could be fully accounted for by the distribution of overdensities within the environments. However, the precision with which this result could be tested was mildly limited by the availability of statistics. When probing only a restricted range of densities within the simulation the number of halos contributing to the mass functions are significantly decreased. This influenced the chosen widths of the restricted density bins in order to limit the statistical noise in the resulting mass functions. Hence, future work may benefit from utilising larger simulations in order to more precisely test the prediction.

This work was able to derive clear conclusions on the influence of the cosmic web on halo mass functions, from within a theoretical framework and as seen empirically within the simulated dataset. All results indicated that scalar halo properties are not heavily influenced by the tidal field beyond the local overdensity.

6.2 The observed cosmic web

Chapter 3 introduced the GAMA spectroscopic redshift survey and its galaxy and group catalogues. It presented a thorough analysis of the impact of various limitations of observational data on the classification of the cosmic web through the tidal tensor prescription, and culminated in a classification of the geometric environments of GAMA galaxies and groups out to $z \sim 0.26$.

It was shown how the pseudo-potential derived from the galaxy density field can be used to estimate the tidal tensor, allowing the tidal tensor prescription to be applied to observational datasets. The variation of geometric

classifications with redshift was tested and shown to differ between low and high redshifts. By estimating the variation of bias and growth factor with redshift, a redshift dependent threshold was proposed. Its implementation showed some improvements, however, it was argued that the benefits did not outweigh the uncertainty in the correct choice of $\lambda_{\rm th}(z)$ and the additional complications involved in its implementation. For this reason the classification of the cosmic web within the GAMA survey was restricted to the lower redshift regime, $z \lesssim 0.26$, and a constant threshold was used.

With further investigation, it is hoped that a robust approximation for a redshift dependent threshold can be developed. This would allow the full survey out to $z \sim 0.5$ to be classified, and potentially see this tidal tensor prescription applied to higher redshift surveys. Not only would this provide significantly more objects for scientific analyses, it also presents an opportunity to analyse the statistics of the cosmic web itself as a function of redshift.

The main difficulties of using observational rather than simulated data to define the web are due to the survey geometry, redshift-space distortions, and the lack of direct knowledge of the underlying density field. Chapter 3 utilised the previous work with the MDR1 simulation in order to quantitatively measure the impact of these effects on the resulting geometric environment classifications. A technique of 'reflecting' galaxies along the survey boundaries was proposed and shown to reduce the impact of the survey geometry. These investigations indicated the combined effect of these three causes of error altered the classifications for $\lesssim 25\%$ of the volume. This error is difficult to reduce as galaxy surveys are not able to measure true real-space coordinates or directly observe the dark matter field. The main source of improvement is likely to come with large wide-field surveys which sample a large enough volume to reduce the edge effects caused by the survey geometry.

These limitations not withstanding, the work documented in this chapter was able to extract the components of the cosmic web from the GAMA survey and provide geometric environment classifications for GAMA galaxies and groups out to $z \sim 0.26$, making the investigations of the remaining chapters possible.

6.3 The luminosity function within the cosmic web

Section 4.1 utilised the classification of the GAMA galaxies presented in Chapter 3 to measure the modulation of the galaxy luminosity function by geometric environment. Metuki et al. (2015) studied the cosmic web within simulations and found a strong dependence of galaxy properties, including stellar mass functions and star formation rate, on web-environment. This suggests that the tidal field plays an as yet poorly understood role in galaxy formation and evolution. The work of this chapter was intended to test this suggestion within observational data by searching for a dependence of the galaxy luminosity function on location within the cosmic web.

A strong variation between different web components was found. Fitting Schechter functions to the resulting luminosity functions allowed this variation to be quantified; the normalisation was seen to increase by a factor of ~ 10 from voids to knots. Although the shapes of the luminosity function did not significantly differ between environments it was found that the turnover point, characterised by M^* in a Schechter function, brightens from voids to knots by ~ 0.5 mag.

One of the limitations of work such as Metuki et al. (2015) is that the effects of the web are not always disentangled from the indirect effects of the local density field. That is, whether the variation of local densities sampled by the different components of the web can account for the variation in the galaxy properties is not tested. In this work that question was addressed by 'shuffling' galaxies between environments based on their local densities. The resulting shuffled catalogues display identical overdensity distributions whilst having information on the geometric environments removed. However, one limitation of this approach is that it requires a volume-limited sample in order to allow the random shuffling of galaxies across the full redshift range. This shuffling across redshift may also affect the results by producing shuffled catalogues with a different redshift distribution to the original catalogues. Nevertheless, it was shown that these shuffled catalogues produce luminosity functions fully consistent with the original luminosity functions extracted from the components of the cosmic web, indicating that the galaxy luminosity function is independent of geometry at a given local

density.

It would however be premature to argue on this basis that the tidal effects have no impact. It is possible that higher-order influences exist but were not detectable at the resolution of this analysis. The range of scales that can be probed is restricted by the requirement of robustly classifying the entire volume of a survey with a limited number density of galaxies. Future work might benefit from studying higher-density regions in higher spatial resolution. This work also focused on only two parameter sets, $\sigma = 4$ and $10\,h^{-1}$Mpc, with a threshold chosen so as to optimise the resulting statistics across all environments. It would be interesting to test how the measured variations in luminosity function differ when a range of parameters are used to define the cosmic web. If optimising the choice of threshold based on the available data was not necessary, one may be tempted to look at 'extreme-environments' and investigate if this lack of a direct dependence persists. Whilst this section considered only the luminosity of galaxies, there are a number of other observable properties that would benefit from a thorough analysis of the direct influence of the cosmic web. Galaxy colour, mass, morphology and star formation history may all exhibit some environmental dependencies and should ideally be investigated in future work. Additionally, there may be reason to expect satellite, or low-mass galaxies to be more strongly linked to their geometrical environment. Further work may thus benefit from investigating these dependencies within different galaxy samples.

In summary, the results of this section were consistent with a picture in which scalar properties of halo and galaxy populations have no direct dependence on their location within the cosmic web. It was shown that whilst a strong variation of the luminosity function on location within the cosmic web was found, all modulation could be accounted for by the indirect dependence on local overdensity.

6.4 The group mass function within the cosmic web

In Section 4.2 the dependence of the group mass function on location within the cosmic web was investigated. This allowed for the fundamental prediction of Chapter 2 – that the halo mass function is independent of geometry for a given

local overdensity – to be tested within observational data.

A significant limitation of this analysis was due to the uncertainty in the mass estimates of groups, which restricts the possibility of this analysis to detect potentially small modulations of the mass function by geometric environment. Group masses are known to be difficult to measure from galaxy surveys, as was evident when comparing three different mass estimates available for the GAMA groups. A puzzlingly large discrepancy was seen between the mass estimates, with the dynamical masses estimated by the GAMA collaboration appearing significantly overestimated when comparing to mass estimates calibrated from weak lensing.

This investigation would benefit from further work on theoretically modelling the global mass function. It was shown how the effect of mass uncertainties can be simulated by convolving the theoretical mass function with a Gaussian. The incompleteness at low masses could also be incorporated into the theoretical mass function. It may be hoped to find a theoretical model in good agreement with the observed mass function with further work.

Conditional group mass functions were extracted from the geometric environments presented in Chapter 3. These were seen to be similar at low masses but diverged for high masses ($\log_{10}(\mathrm{M}/h^{-1}\mathrm{M}_{\odot}) \gtrsim 13$). To search for any direct dependence on geometry, mass functions were extracted from regions with local overdensities within a specific range. This approach was limited by the number of groups in the survey and large Poisson errors on the number counts were seen in some cases. However, in general, the restricted-density mass functions from different components of the cosmic web were consistent with each other.

Section 4.2 also presented results from a second approach taken to disentangling the effects of the local density field from the tidal field. Groups were shuffled between environments based on their local densities, as in Section 4.1. However, in this case the density used was the smoothed overdensity of the $(3\,h^{-1}\mathrm{Mpc})^3$ cell in which the group resides. Again, the shuffling across redshift may affect the results by producing shuffled catalogues with a different redshift distribution to the original catalogue. This may have a significant effect here as in this case the sample was not volume-limited but contains strong redshift dependencies. Extensions of this work may benefit from incorporating the density measurements of Section 4.1, derived from the number of galaxies within a sphere of variable

radius. This would allow the shuffling to be driven by arguably more accurate measurements of local density that are able to probe a range of scales. The shuffling approach again produced mostly consistent mass functions, indicating that the group mass function is independent of geometry for a given local density. It was difficult to ascertain whether the small deviations seen were statistically consistent due to the lack of reliable uncertainties on the mass functions. Hence, further work in this regard would strengthen the conclusions which can be drawn.

This section would also benefit from further work utilising mock catalogues. As well as providing more groups, reducing statistical uncertainties, mock catalogues would also provide insights into the uncertainties on the recovered mass functions. It would also be interesting to test whether the mass functions extracted from mock catalogues, of a given cosmology, are in agreement with the results presented here. Additionally, the variation of the mass functions with redshift has not yet been studied in detail. It is possible that the effects of the geometric environments are not constant over redshift, and, particularly when looking at restricted overdensity regions or when shuffling the catalogues, the effect of the altered redshift distributions of groups should receive further attention.

This section showed how the group mass function varies within different regions of the cosmic web. No strong evidence of a direct dependence of the group mass function on geometric environment was found. The results of this section suggest that the variations of the group mass function can be accounted for by the different densities sampled by the environments. These results are consistent with the prediction of Chapter 2 and, again, with a picture in which scalar properties of halo and galaxy populations have no direct dependence on their location within the cosmic web.

6.5 Stellar assembly histories within the cosmic web

The final chapter of this thesis investigated the stellar assembly histories of GAMA galaxies within different components of the cosmic web. This was motivated by the need to look beyond the 'final state' properties that had been studied in previous chapters. All previous results presented in this thesis

indicated that scalar properties of LSS are not strongly influenced by their web-environment. For this reason it was desirable to look beyond the final state properties such as mass and luminosity and investigate the *histories* of LSS. A positive detection of a dependence on geometric environment would provide a strong indication of assembly bias, potentially leading to improved halo occupation statistic formalisms and providing a better way to re-parameterise galaxy-halo relation models.

A full spectral fitting code, VESPA, was used to recover star formation histories for the GAMA galaxies. This analysis is heavily reliant on the ability of VESPA to robustly recover the histories, which is dependent on both the accuracy of the modelling and the quality of the spectroscopic data. On an individual basis, the star formation history of a spectrum can only be recovered with a modest level of confidence. However, the large number of galaxies within the GAMA catalogue allowed many results to be stacked in order to reduce the uncertainties.

GAMA spectra are known to suffer from imperfect photometric calibration, particularly in the blue end. For this reason a scaling algorithm was produced in order to calibrate the spectra to the SDSS magnitudes based on linear interpolation. This was shown to remove mostly stochastic offsets and lead to more physical star formation histories on average. The scaling applied was a relatively simple approach, given the importance of photometric calibration, this investigation and many others would benefit from further work ensuring that the calibration is correct or testing more complex approaches to the calibration. By comparing the stacked spectra of duplicate objects observed by both SDSS and GAMA, it was shown that a significant discrepancy remains, with the spectra displaying different continuum shapes on average. This discrepancy should be explored further in order to ascertain its origin.

The full star formation histories displayed an unexpected upturn at recent ages which was seen to be affected by the inclusion of photometry and hence the detailed reliability of the SFHs at recent times was questioned. For this reason, three statistics were introduced to characterise the formation time of the galaxies. It was shown that these statistics often displayed a strong bimodality and it was found that this bimodality was a feature of the blue, dusty or low signal to noise galaxy populations in particular. The strong correlation of the bimodality with galaxy colour indicates it may be physically driven, though the extent to which it

is due to an inability of VESPA to accurately recover star formation at moderate ages remains to be fully investigated.

The average values of these statistics within different components of the web were calculated. It was found that the fraction of young stars decreases between voids and knots, consistent with the view that void galaxies display the most recent star formation activity. It was also seen that, when using alternative statistics to characterise formation time (t_{85} and mass weighted age), the void galaxies appeared older than knot galaxies. These two results are not necessarily inconsistent due to the different statistics used but suggests an interesting result whereby void galaxies, for example, may be considered simultaneously young and old.

To search for assembly bias these average statistics were computed again for galaxies residing in halos of similar mass. This removed any clear dependence on geometric environment of the fraction of young stars, consistent with no assembly bias. However, the other two statistics retained some dependence, indicating that assembly bias may exist. In order to ascertain whether these results are consistent with no assembly bias, or if they indicate assembly bias with a strong statistical significance, requires further investigation into the uncertainties associated with the results.

This chapter details an ongoing project and hence there is much scope for further work. Carrying out tests on simulated spectra with known star formation histories is of a high priority. This will indicate the robustness of the SFHs recovered through VESPA and shed light on the cause of the bimodalities seen. It also presents an opportunity to better optimise the statistics used to indicate formation time and investigate how well the currently used statistics perform.

There is also a need to further investigate various technicalities, such as the effect of the choice of stellar population models and of the dust models on the recovered SFHs. The differing redshift distributions of galaxies from different geometric environments also requires attention. Further investigation may benefit from using a weighted sample to correct for these differing distributions, by looking at specific redshift ranges, or by altering the definition of the statistics, for example, to only consider star formation that occurred before a certain lookback time.

With further investigation the work of this chapter stands to provide exciting

insights into the role of the environment during star formation. The initial results indicate that there may be some evidence of assembly bias within the galaxy population. Even if all results were found to be consistent with no assembly bias on further investigation, this would still stand as an important result. The quantification of formation times within observational datasets such as GAMA addresses a difficult and not often attempted task with the potential to extract valuable information on the influence of large-scale environments on galaxy formation.

6.6 Final remarks

This thesis has investigated the modulating effects of the cosmic web on LSS. Driven by the fundamental prediction of Chapter 2, this was undertaken with a critical eye as to the extent to which any modulations can be considered direct evidence of the influence of tidal forces, or whether they are due to correlations with other properties such as the local density.

Throughout all investigations no strong evidence of a direct impact of the web environment on LSS was seen. Hence, this thesis provides strong support for the assertion that scalar properties of halo and galaxy populations have no direct dependence on their location within the cosmic web. This is not at odds with the increasing body of evidence that tensor properties are influenced by the tidal field; this remains an active field of study in order to better understand and quantify these effects. Whilst this thesis focused on quantities such as mass and luminosity, there are many further properties of LSS which may exhibit a correlation and hence warrant further investigation. For example, there is the interesting question of whether properties such as the histories of galaxies display a dependence on geometric environment not at odds with their geometry-independent final states.

Hence, further investigation is required in order to argue with confidence that tidal forces have no impact. For example, by probing a larger range of scales; this may be made possible by future surveys which are able to detect fainter galaxies producing a higher number density galaxy catalogue. More accurate measurements of group masses would also allow tighter constraints on the dependence of the group mass function on geometric environment.

Furthermore, this thesis considers only one definition of the cosmic web whilst a number of different approaches have been proposed. Other classifications may find dependencies not detected within the work presented here. The two free parameters of the tidal tensor prescription employed here were chosen, in part, based on the limited availability of data. Future investigations considering the 'extremes of environment', for example voids defined with a substantially lower $\lambda_{\rm th}$, may find evidence of modulating effects of these environments on large-scale structure.

Bibliography

[1] Aarseth S. J., 1963, MNRAS, 126, 223

[2] Abazajian K., et al., 2003, The Astronomical Journal, 126, 2081

[3] Abel T., Hahn O., Kaehler R., 2012, MNRAS, 427, 61

[4] Abell G. O., 1965, ARA&A, 3, 1

[5] Ade et al., 2014, A&A, 571, A1

[6] Alcock C., Paczynski B., 1979, Nature, 281, 358

[7] Alonso D., Eardley E., Peacock J. A., 2015, MNRAS, 447, 2683

[8] Alpaslan M., et al., 2014, MNRAS, 438, 177

[9] Aragon-Calvo M. A., 2012, ArXiv e-prints

[10] Aragón-Calvo M. A., Jones B. J. T., van de Weygaert R., van der Hulst J. M., 2007, A&A, 474, 315

[11] Aragon-Calvo M. A., Shandarin S. F., Szalay A., 2010, ArXiv e-prints

[12] Aragón-Calvo M. A., van de Weygaert R., Jones B. J. T., 2010, MNRAS, 408, 2163

[13] Arnold V. I., Shandarin S. F., Zeldovich I. B., 1982, Geophysical and Astrophysical Fluid Dynamics, 20, 111

[14] Avila-Reese V., Colín P., Gottlöber S., Firmani C., Maulbetsch C., 2005, ApJ, 634, 51

[15] Baldry I. K., et al., 2010, MNRAS, 404, 86

[16] Balogh M. L., Baldry I. K., Nichol R., Miller C., Bower R., Glazebrook K., 2004, ApJ, 615, L101

[17] Bardeen J. M., Bond J. R., Kaiser N., Szalay A. S., 1986, ApJ, 304, 15

[18] Barrow J. D., Bhavsar S. P., Sonoda D. H., 1985, MNRAS, 216, 17

[19] Berlind A. A., Weinberg D. H., 2002, ApJ, 575, 587

[20] Berlind A. A., Weinberg D. H., Benson A. J., Baugh C. M., Cole S., Davé R., Frenk C. S., Jenkins A., Katz N., Lacey C. G., 2003, ApJ, 593, 1

[21] Bernardeau F., 1994, ApJ, 427, 51

[22] Bernardeau F., Colombi S., Gaztañaga E., Scoccimarro R., 2002, Physics Reports, 367, 1

[23] Bertschinger E., 1998, ARA&A, 36, 599

[24] Blake C., et al., 2012, MNRAS, 425, 405

[25] Blanton M. R., Berlind A. A., 2007, ApJ, 664, 791

[26] Blanton M. R., Eisenstein D., Hogg D. W., Schlegel D. J., Brinkmann J., 2005, ApJ, 629, 143

[27] Bond J. R., Cole S., Efstathiou G., Kaiser N., 1991, ApJ, 379, 440

[28] Bond J. R., Myers S. T., 1996, ApJS, 103, 1

[29] Bower R. G., 1991, MNRAS, 248, 332

[30] Bower R. G., et al., 2006, MNRAS, 370, 645

[31] Brainerd T. G., Scherrer R. J., Villumsen J. V., 1993, ApJ, 418, 570

[32] Brough S., et al., 2013, MNRAS, 435, 2903

[33] Cabré A., Gaztañaga E., 2009, MNRAS, 393, 1183

[34] Carroll S. M., Press W. H., Turner E. L., 1992, ARA&A, 30, 499

[35] Cautun M., van de Weygaert R., Jones B. J. T., 2013, MNRAS, 429, 1286

[36] Codis S., Pichon C., Devriendt J., Slyz A., Pogosyan D., Dubois Y., Sousbie T., 2012, MNRAS, 427, 3320

[37] Colberg J. M., et al., 2008, MNRAS, 387, 933

[38] Cole S., 2011, MNRAS, 416, 739

[39] Cole S., Helly J., Frenk C. S., Parkinson H., 2008, MNRAS, 383, 546

[40] Cole S., Kaiser N., 1989, MNRAS, 237, 1127

[41] Coles P., Jones B., 1991, MNRAS, 248, 1

[42] Colless M., et al., 2001, MNRAS, 328, 1039

[43] Colless M., et al., 2003a, VizieR Online Data Catalog, 7226, 0

[44] Colless M., et al., 2003b, ArXiv Astrophysics e-prints

[45] Colombi S., Pogosyan D., Souradeep T., 2000, Physical Review Letters, 85, 5515

[46] Contreras S., Baugh C. M., Norberg P., Padilla N., 2015, ArXiv e-prints

[47] Cooray A., Sheth R., 2002, Physics Reports, 372, 1

[48] Croton D. J., et al., 2005, MNRAS, 356, 1155

[49] Darvish B., Sobral D., Mobasher B., Scoville N. Z., Best P., Sales L. V., Smail I., 2014, ApJ, 796, 51

[50] Davis M., Efstathiou G., Frenk C. S., White S. D. M., 1985, ApJ, 292, 371

[51] de Bernardis P., et al., 2000, Nature, 404, 955

[52] de Lapparent V., Geller M. J., Huchra J. P., 1986, ApJ, 302, L1

[53] Dekel A., 1994, ARA&A, 32, 371

[54] Dodelson S., 2003, Modern cosmology. Academic Press

[55] Doroshkevich A., Tucker D. L., Allam S., Way M. J., 2004, A&A, 418, 7

[56] Doroshkevich A. G., 1970, Astrofizika, 6, 581

[57] Dressler A., 1980, ApJ, 236, 351

[58] Driver S. P., et al., 2009, Astronomy and Geophysics, 50, 12

[59] Driver S. P., et al., 2011, MNRAS, 413, 971

[60] Dubois Y., et al., 2014, MNRAS, 444, 1453

[61] Duffy A. R., Schaye J., Kay S. T., Dalla Vecchia C., Battye R. A., Booth C. M., 2010, MNRAS, 405, 2161

[62] Eardley E., Peacock J. A., McNaught-Roberts T., Heymans C., Norberg P., Alpaslan M., Baldry I., Bland-Hawthorn J., Brough S., Cluver M. E., Driver S. P., Farrow D. J., Liske J., Loveday J., Robotham A. S. G., 2015, MNRAS, 448, 3665

[63] Efstathiou G., Davis M., White S. D. M., Frenk C. S., 1985, ApJS, 57, 241

[64] Efstathiou G., Ellis R. S., Peterson B. A., 1988, MNRAS, 232, 431

[65] Einasto J., Suhhonenko I., Hütsi G., Saar E., Einasto M., Liivamägi L. J., Müller V., Starobinsky A. A., Tago E., Tempel E., 2011, A&A, 534, A128

[66] Eriksen H. K., Novikov D. I., Lilje P. B., Banday A. J., Górski K. M., 2004, ApJ, 612, 64

[67] Falck B. L., Neyrinck M. C., Szalay A. S., 2012, ApJ, 754, 126

[68] Faltenbacher A., White S. D. M., 2010, ApJ, 708, 469

[69] Forero-Romero J. E., Contreras S., Padilla N., 2014, MNRAS, 443, 1090

[70] Forero-Romero J. E., Hoffman Y., Gottlöber S., Klypin A., Yepes G., 2009, MNRAS, 396, 1815

[71] Gao L., Springel V., White S. D. M., 2005, MNRAS, 363, L66

[72] Gao L., White S. D. M., 2007, MNRAS, 377, L5

[73] Garilli B., et al., 2014, A&A, 562, A23

[74] Geller M. J., Huchra J. P., 1983, ApJS, 52, 61

[75] Geller M. J., Huchra J. P., 1989, Science, 246, 897

[76] Goldberg D. M., Vogeley M. S., 2004, ApJ, 605, 1

[77] Gómez P. L., Nichol R. C., Miller C. J., Balogh M. L., Goto T., Zabludoff A. I., Romer A. K., Bernardi M., Sheth R., Hopkins A. M., Castander F. J., Connolly A. J., Schneider D. P., Brinkmann J., Lamb D. Q., SubbaRao M., York D. G., 2003, ApJ, 584, 210

[78] Gottlöber S., Łokas E. L., Klypin A., Hoffman Y., 2003, MNRAS, 344, 715

[79] Gregory S. A., Thompson L. A., 1978, ApJ, 222, 784

[80] Guo Q., Tempel E., Libeskind N. I., 2014, ArXiv: 1403.5563

[81] Guth A. H., 1981, Phys. Rev. D, 23, 347

[82] Guzzo L., et al., 2008, Nature, 451, 541

[83] Hahn O., Porciani C., Carollo C. M., Dekel A., 2007, MNRAS, 375, 489

[84] Hahn O., Porciani C., Dekel A., Carollo C. M., 2009, MNRAS, 398, 1742

[85] Han J., Eke V. R., Frenk C. S., Mandelbaum R., Norberg P., Schneider M. D., Peacock J. A., Jing Y., Baldry I., Bland-Hawthorn J., Brough S., Brown M. J. I., Liske J., Loveday J., Robotham A. S. G., 2015, MNRAS, 446, 1356

[86] Harrison E. R., 1970, Phys. Rev. D, 1, 2726

[87] Hawkins E., et al., 2003, MNRAS, 346, 78

[88] Hayashino T., et al., 2004, The Astronomical Journal, 128, 2073

[89] Hidding J., Shandarin S. F., van de Weygaert R., 2014, MNRAS, 437, 3442

[90] Hoffman Y., Metuki O., Yepes G., Gottlöber S., Forero-Romero J. E., Libeskind N. I., Knebe A., 2012, MNRAS, 425, 2049

[91] Hoffmann K., Bel J., Gaztañaga E., Crocce M., Fosalba P., Castander F. J., 2015, MNRAS, 447, 1724

[92] Hopkins A. M., et al., 2013, MNRAS, 430, 2047

[93] Hütsi G., Einasto J., Tucker D. L., Saar E., Einasto M., Müller V., Heinämäki P., Allam S. S., 2002, arXiv:astro-ph/0212327

[94] Jenkins A., Frenk C. S., White S. D. M., Colberg J. M., Cole S., Evrard A. E., Couchman H. M. P., Yoshida N., 2001, MNRAS, 321, 372

[95] Kaiser N., 1984, ApJ, 284, L9

[96] Kaiser N., 1987, MNRAS, 227, 1

[97] Kauffmann G., White S. D. M., Guiderdoni B., 1993, MNRAS, 264, 201

[98] Kitaura F.-S., Jasche J., Metcalf R. B., 2010, MNRAS, 403, 589

[99] Klypin A., Holtzman J., 1997, ArXiv Astrophysics e-prints

[100] Klypin A. A., Shandarin S. F., 1983, MNRAS, 204, 891

[101] Lacey C., Cole S., 1993, MNRAS, 262, 627

[102] Lacey C., Cole S., 1994, MNRAS, 271, 676

[103] Leclercq F., Jasche J., Wandelt B., 2015, A&A, 576, L17

[104] Lee J., Rey S. C., Kim S., 2014, ApJ, 791, 15

[105] Lemson G., Kauffmann G., 1999, MNRAS, 302, 111

[106] Libeskind N. I., Hoffman Y., Forero-Romero J., Gottlöber S., Knebe A., Steinmetz M., Klypin A., 2013, MNRAS, 428, 2489

[107] Libeskind N. I., Hoffman Y., Knebe A., Steinmetz M., Gottlöber S., Metuki O., Yepes G., 2012, MNRAS, 421, L137

[108] Libeskind N. I., Knebe A., Hoffman Y., Gottloeber S., 2014, ArXiv e-prints

[109] Loeb A., Furlanetto S. R., 2013, The First Galaxies in the Universe. Princeton University Press

[110] Ludlow A. D., Porciani C., 2011, MNRAS, 413, 1961

[111] Lyth D. H., Liddle A. R., 2009, The Primordial Density Perturbation, 1 edn. Cambridge University Press

[112] Maggiore M., Riotto A., 2010, ApJ, 711, 907

[113] Magliocchetti M., Porciani C., 2003, MNRAS, 346, 186

[114] Maraston C., 2005, MNRAS, 362, 799

[115] Martino M. C., Sheth R. K., 2009, MNRAS, 394, 2109

[116] McNaught-Roberts T., et al., 2014, Astrophysics

[117] Mecke K. R., Buchert T., Wagner H., 1994, A&A, 288, 697

[118] Melott A. L., Pellman T. F., Shandarin S. F., 1994, MNRAS, 269, 626

[119] Metuki O., Libeskind N. I., Hoffman Y., Crain R. A., Theuns T., 2015, MNRAS, 446, 1458

[120] Miller A. D., Caldwell R., Devlin M. J., Dorwart W. B., Herbig T., Nolta M. R., Page L. A., Puchalla J., Torbet E., Tran H. T., 1999, ApJ, 524, L1

[121] Mo H. J., Jing Y. P., White S. D. M., 1996, MNRAS, 282, 1096

[122] Mo H. J., White S. D. M., 1996, MNRAS, 282, 347

[123] Moster B. P., Somerville R. S., Maulbetsch C., van den Bosch F. C., Macciò A. V., Naab T., Oser L., 2010, ApJ, 710, 903

[124] Muldrew S. I., et al., 2012, MNRAS, 419, 2670

[125] Musso M., Paranjape A., Sheth R. K., 2012, MNRAS, 427, 3145

[126] Norberg P., et al., 2001, MNRAS, 328, 64

[127] Novikov D., Colombi S., Doré O., 2006, MNRAS, 366, 1201

[128] Nusser A., Dekel A., 1992, ApJ, 391, 443

[129] Okada H., Totani T., Tsujikawa S., 2013, Phys. Rev. D, 87, 103002

[130] Olive K. A., Skillman E. D., 2004, ApJ, 617, 29

[131] Padmanabhan T., 2002, Theoretical Astrophysics, Volume III: Galaxies and Cosmology. Cambridge University Press

[132] Paranjape A., Sheth R. K., Desjacques V., 2013, MNRAS, 431, 1503

[133] Peacock J., 2001, `www.roe.ac.uk/~jap/teaching/cos4notes/notes.pdf`

[134] Peacock J. A., 2007, MNRAS, 379, 1067

[135] Peacock J. A., Heavens A. F., 1990, MNRAS, 243, 133

[136] Peacock J. A., Smith R. E., 2000, MNRAS, 318, 1144

[137] Peebles J. P. E., 1993, Principles of Physical Cosmology. Princeton University Press

[138] Peebles P. J. E., 1980, The Large Scale Structure of the Universe, 1 edn. Princeton University Press

[139] Peebles P. J. E., Groth E. J., 1975, ApJ, 196, 1

[140] Peebles P. J. E., Silk J., 1990, Nature, 346, 233

[141] Penzias A. A., Wilson R. W., 1965, ApJ, 142, 419

[142] Percival W. J., et al., 2004, MNRAS, 353, 1201

[143] Percival W. J., Scott D., Peacock J. A., Dunlop J. S., 2003, MNRAS, 338, L31

[144] Petrosian V., 1976, ApJ, 209, L1

[145] Planck Collaboration Ade P. A. R., Aghanim N., Arnaud M., Ashdown M., Aumont J., Baccigalupi C., Banday A. J., Barreiro R. B., Bartlett J. G., et al. 2015, ArXiv e-prints: 1502.01589

[146] Platen E., et al., 2007, MNRAS, 380, 551

[147] Prada F., Klypin A. A., Cuesta A. J., Betancort-Rijo J. E., Primack J., 2012, MNRAS, 423, 3018

[148] Press W. H., Schechter P., 1974, ApJ, 187, 425

[149] Reed D. S., Bower R., Frenk C. S., Jenkins A., Theuns T., 2007, MNRAS, 374, 2

[150] Reid B. A., et al., 2012, MNRAS, 426, 2719

[151] Robotham A., et al., 2010, PASA, 27, 76

[152] Robotham A. S. G., et al., 2011, MNRAS, 416, 2640

[153] Robotham A. S. G., et al., 2013, MNRAS, 431, 167

[154] Rossi G., 2013, MNRAS, 430, 1486

[155] Samushia L., Percival W. J., Raccanelli A., 2012, MNRAS, 420, 2102

[156] Schechter P., 1976, ApJ, 203, 297

[157] Scoccimarro R., 1998, MNRAS, 299, 1097

[158] Shandarin S., Habib S., Heitmann K., 2012, Phys. Rev. D, 85, 083005

[159] Shandarin S. F., 2011, JCAP, 5, 15

[160] Sheth R. K., Mo H. J., Tormen G., 2001, MNRAS, 323, 1

[161] Sheth R. K., Tormen G., 1999, MNRAS, 308, 119

[162] Sousbie T., 2011, MNRAS, 414, 350

[163] Sousbie T., Pichon C., Colombi S., Novikov D., Pogosyan D., 2008, MNRAS, 383, 1655

[164] Sousbie T., Pichon C., Courtois H., Colombi S., Novikov D., 2008, ApJ, 672, L1

[165] Sousbie T., Pichon C., Kawahara H., 2011, MNRAS, 414, 384

[166] Springel V., White S. D. M., Jenkins A., Frenk C. S., Yoshida N., Gao L., Navarro J., Thacker R., Croton D., Helly J., Peacock J. A., Cole S., Thomas P., Couchman H., Evrard A., Colberg J., Pearce F., 2005, Nature, 435, 629

[167] Steinmetz M., Navarro J. F., 2002, Nature, 7, 155

[168] Stoica R. S., Martínez V. J., Saar E., 2010, A&A, 510, A38

[169] Sutter P. M., Pisani A., Wandelt B. D., Weinberg D. H., 2014, MNRAS, 443, 2983

[170] Tempel E., Libeskind N. I., Hoffman Y., Liivamägi L. J., Tamm A., 2014, MNRAS, 437, L11

[171] Tempel E., Saar E., Liivamägi L. J., Tamm A., Einasto J., Einasto M., Müller V., 2011, A&A, 529, A53

[172] Tinker J., Kravtsov A. V., Klypin A., Abazajian K., Warren M., Yepes G., Gottlöber S., Holz D. E., 2008, ApJ, 688, 709

[173] Tinker J. L., Conroy C., Norberg P., Patiri S. G., Weinberg D. H., Warren M. S., 2008, ApJ, 686, 53

[174] Tojeiro R., Heavens A. F., Jimenez R., Panter B., 2007, MNRAS, 381, 1252

[175] Tojeiro R., Wilkins S., Heavens A. F., Panter B., Jimenez R., 2009, ApJS, 185, 1

[176] Vale A., Ostriker J. P., 2004, MNRAS, 353, 189

[177] van Albada T. S., Bahcall J. N., Begeman K., Sancisi R., 1985, ApJ, 295, 305

[178] van den Bosch F. C., Yang X., Mo H. J., Weinmann S. M., Macciò A. V., More S., Cacciato M., Skibba R., Kang X., 2007, MNRAS, 376, 841

[179] von Hoerner S., 1960, Zeitschrift fr Astrophysik, 50, 184

[180] Wagner C., Verde L., Boubekeur L., 2010, JCAP, 10, 22

[181] Wang L., Weinmann S. M., De Lucia G., Yang X., 2013, MNRAS, 433, 515

[182] Warren M. S., Quinn P. J., Salmon J. K., Zurek W. H., 1992, ApJ, 399, 405

[183] Watson W. A., Iliev I. T., D'Aloisio A., Knebe A., Shapiro P. R., Yepes G., 2013, MNRAS, 433, 1230

[184] Wechsler R. H., Zentner A. R., Bullock J. S., Kravtsov A. V., Allgood B., 2006, ApJ, 652, 71

[185] Weinberg S., 1972, Gravitation and Cosmology: Principles and Applications of the General Theory of Relativity. John Wiley & Sons, Inc.

[186] White S. D. M., Davis M., Efstathiou G., Frenk C. S., 1987, Nature, 330, 451

[187] Wijesinghe D. B., et al., 2012, MNRAS, 423, 3679

[188] Wu K. K. S., Lahav O., Rees M. J., 1999, Nature, 397, 225

[189] Yadav J., Bharadwaj S., Pandey B., Seshadri T. R., 2005, MNRAS, 364, 601

[190] Yan H., Fan Z., White S. D. M., 2012, ArXiv: 1203.1225

[191] Yan H., Fan Z., White S. D. M., 2013, MNRAS, 430, 3432

[192] Yang X., Mo H. J., van den Bosch F. C., 2003, MNRAS, 339, 1057

[193] Yang X., Mo H. J., van den Bosch F. C., 2006, ApJ, 638, L55

[194] York D. G., et al., 2000, The Astronomical Journal, 120, 1579

[195] Yoshisato A., Morikawa M., Gouda N., Mouri H., 2006, ApJ, 637, 555

[196] Zel'dovich Y. B., 1970, AAP, 5, 84

[197] Zeldovich Y. B., 1972, MNRAS, 160, 1P

[198] Zentner A. R., Hearin A. P., van den Bosch F. C., 2014, MNRAS, 443, 3044

[199] Zhang Y., Yang X., Faltenbacher A., Springel V., Lin W., Wang H., 2009, ApJ, 706, 747

[200] Zhang Y., Yang X., Wang H., Wang L., Mo H. J., van den Bosch F. C., 2013, ApJ, 779, 160

[201] Zheng Z., Berlind A. A., Weinberg D. H., Benson A. J., Baugh C. M., Cole S., Davé R., Frenk C. S., Katz N., Lacey C. G., 2005, ApJ, 633, 791

Appendices

Appendix A

A.1 Correlations between environment and density for Gaussian fields

A.1.1 The eigenvalue distribution

Consider a Gaussian potential field $\tilde{\phi}$ smoothed over a length scale R. Since the tidal tensor $\hat{T}$ is a symmetric matrix only 6 of its components are independent. Here they will be labelled with a single index: $T_A = (T_{11}, T_{22}, T_{33}, T_{23}, T_{31}, T_{12})$. It is straightforward to calculate the covariance matrix of the $T_A{}'s$:

$$\langle T_A T_B \rangle = \frac{\sigma_R^2}{15} \begin{pmatrix} 3 & 1 & 1 & 0 & 0 & 0 \\ 1 & 3 & 1 & 0 & 0 & 0 \\ 1 & 1 & 3 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \end{pmatrix}, \tag{A.1}$$

where σ_R^2 is given in Equation (2.17) with $R_a = R_b = R$. This matrix can be diagonalized by changing to the variables:

$$\begin{aligned} &\tau_1 \equiv \nu \equiv \frac{1}{\sigma_R}(T_1 + T_2 + T_3), \;\; \tau_2 \equiv \rho \equiv \frac{1}{2\sigma_R}(T_1 - T_3), \\ &\tau_3 \equiv \theta \equiv \frac{1}{2\sigma_R}(T_1 - 2T_2 + T_3), \\ &\tau_4 \equiv \frac{T_4}{\sigma_R}, \;\; \tau_5 \equiv \frac{T_5}{\sigma_R}, \;\; \tau_6 \equiv \frac{T_6}{\sigma_R}. \end{aligned} \tag{A.2}$$

Notice that by definition, ν is proportional to the local density contrast: $\delta = \nu\sigma_R$ and that ρ and θ are trivially related to the ellipticity $e \equiv \rho/\nu$ and prolateness $p \equiv \theta/\nu$. In

terms of these new variables the covariance matrix is diagonal:

$$\mu_{AB} \equiv \langle \tau_A \tau_B \rangle = \mathrm{diag}\left(1, \frac{1}{15}, \frac{1}{5}, \frac{1}{15}, \frac{1}{15}, \frac{1}{15}\right) \tag{A.3}$$

and the joint distribution of the T_A's is

$$P(\{T_A\}) \prod_A dT_A = \frac{e^{-Q/2}}{\sqrt{(2\pi)^6 \det(\hat{\mu})}} \prod_A d\tau_A,$$
$$Q \equiv \nu^2 + 15\rho^2 + 5\theta^2 + 15(\tau_4^2 + \tau_5^2 + \tau_6^2). \tag{A.4}$$

This holds in any coordinate system, but it will be most useful in the one in which $\hat{T}$ is diagonal (i.e. $\hat{T} = \mathrm{diag}(\lambda_1, \lambda_2, \lambda_3, 0, 0, 0)$). As proved by Bardeen et al. (1986) the volume element of the space of 3×3 symmetric matrices can be written in terms of the matrix eigenvalues and the Euler angles of the rotation necessary to diagonalise it:

$$\prod_A dT_A = |(\lambda_1 - \lambda_2)(\lambda_2 - \lambda_3)(\lambda_1 - \lambda_3)|\, d\lambda_1\, d\lambda_2\, d\lambda_3 \frac{d\Omega_3}{6}, \tag{A.5}$$

where $d\Omega_3$ is the volume element of $\mathcal{S}^3$ (the total volume of which is $2\pi^2$). Up to now no specific ordering for the eigenvalues has been chosen. There are 6 possible orderings, and the probability density is symmetric with respect to these, therefore imposing a specific ordering would introduce a factor 6 in the probability density above. Choosing $\lambda_1 \geq \lambda_2 \geq \lambda_3$ (which implies $\theta \in [-\rho, \rho]$, $\rho \in [0, \infty)$) and integrating out the irrelevant angular part produces the probability distribution given in Equation (2.8).

A.1.2 Correlation with the local density contrast

Now let us consider the distribution of the density contrast smoothed over a scale R_h in different regions classified via the tidal tensor eigenvalues smoothed over a scale R_e. Here we will use the notation introduced in Section 2.3, with $\delta_e \equiv \mathrm{Tr}(\hat{T})$.

The covariance of δ_h with the T_A's is easy to calculate:

$$\langle \delta_h\, T_A \rangle = \frac{\sigma_{eh}^2}{3}(\delta_{A1} + \delta_{A2} + \delta_{A3}) \tag{A.6}$$

Defining $G_a = (\delta_h, T_A)$, $(a = 0, ..., 6)$ the full covariance matrix is

$$\langle G_a G_b \rangle = \frac{1}{15} \times \begin{pmatrix} 15\sigma_{hh}^2 & 5\sigma_{eh}^2 & 5\sigma_{eh}^2 & 5\sigma_{eh}^2 & 0 & 0 & 0 \\ 5\sigma_{eh}^2 & 3\sigma_{ee}^2 & \sigma_{ee}^2 & \sigma_{ee}^2 & 0 & 0 & 0 \\ 5\sigma_{eh}^2 & \sigma_{ee}^2 & 3\sigma_{ee}^2 & \sigma_{ee}^2 & 0 & 0 & 0 \\ 5\sigma_{eh}^2 & \sigma_{ee}^2 & \sigma_{ee}^2 & 3\sigma_{ee}^2 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & \sigma_{ee}^2 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & \sigma_{ee}^2 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & \sigma_{ee}^2 \end{pmatrix}. \tag{A.7}$$

As before, in order to simplify this matrix, define:

$$\begin{aligned} &\gamma_0 \equiv \nu_h \equiv \frac{\delta_h}{\sigma_{hh}}, \;\; \gamma_1 \equiv \nu_e \equiv \frac{T_1 + T_2 + T_3}{\sigma_{ee}}, \\ &\gamma_2 \equiv \rho \equiv \frac{T_1 - T_3}{2\sigma_{ee}}, \;\; \gamma_3 \equiv \theta \equiv \frac{T_1 - 2T_2 + T_3}{2\sigma_{ee}}, \\ &\gamma_4 \equiv \frac{T_4}{\sigma_{ee}}, \;\; \gamma_5 \equiv \frac{T_5}{\sigma_{ee}}, \;\; \gamma_6 \equiv \frac{T_6}{\sigma_{ee}}, \end{aligned} \tag{A.8}$$

the covariance of which is

$$\langle \gamma_a \gamma_b \rangle = \begin{pmatrix} 1 & \varepsilon & 0 & 0 & 0 & 0 & 0 \\ \varepsilon & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1/15 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1/5 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1/15 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1/15 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1/15 \end{pmatrix}, \tag{A.9}$$

Now the same procedure as before can be followed: change the volume element of the T_A's to the one in the space of eigenvalues and rotations, change to the coordinate system in which $\hat{T}$ is diagonal, transform everything to our variables $\{\gamma_a\}$, choose a specific ordering for the eigenvalues and integrate out the irrelevant angular part. At the end of the day the distribution given in Equation (2.26) is obtained.

A.2 The effective universe approach

Consider a spherical perturbation in an otherwise homogeneous universe. It is a well-known result in gravitational theory that at any distance from the centre of the perturbation, it must evolve as a parallel FRW cosmology with some effective cosmological parameters which can be entirely determined in terms of the amplitude of the density perturbation. This result allows us to interpret the conditional mass function for an environment with overdensity δ_e as the mass function in the corresponding effective universe.

The 'environmental' cosmological parameters are related to the background ones and the perturbation's overdensity through:

$$\Omega_M^e = \Omega_M^{\rm BG}\,(1+\delta)/\eta^2 \quad \Omega_\Lambda^e = \Omega_\Lambda^{\rm BG}/\eta^2 \quad H_0^e = \eta^2\, H_0^{\rm BG}, \tag{A.10}$$

where the superscripts BG and e denote quantities in the background and in the effective universe respectively. The ratio between the current expansion rates inside and outside the perturbation, η, can be fixed by imposing that the age of the Universe

$$t_{\rm BB} = \frac{1}{H_0}\int_0^1 \frac{dx}{x\sqrt{\Omega_M x^{-3} + \Omega_\Lambda + \Omega_k x^{-2}}} \tag{A.11}$$

must be the same as measured by any observer. This effectively implies that the perturbation must be a purely growing mode that disappears at early times.

Once the effective cosmological parameters are known, the scaling factor D_g in Equation (2.21) is given by the ratio of the growth factors in the two cosmologies. Normalising this ratio to be 1 at early times (where the perturbation gradually disappears), this quantity is given by

$$D_g = \left(\frac{\Omega_M^{\rm BG} h_{\rm BG}^2}{\Omega_M^e h_e^2}\right)^{1/3} \frac{\Omega_M^e}{\Omega_M^{\rm BG}} \frac{g(\Omega_M^e, \Omega_\Lambda^e)}{g(\Omega_M^{\rm BG}, \Omega_\Lambda^{\rm BG})}, \tag{A.12}$$

where

$$g(\Omega_M, \Omega_\Lambda) = \int_0^1 dx \left(\frac{x}{\Omega_M + \Omega_\Lambda x^3 + \Omega_k x}\right)^{3/2}. \tag{A.13}$$

Notice that at this point we have not taken into account the size of the environment. For large smoothing scales or comparatively small halo masses, this is not an important concern: we may treat the environment as an infinite effective universe in which halos of any mass may form. In practice however, the mass of the largest halos ($M \sim 10^{15}\, M_\odot \rightarrow R_h \sim 15$Mpc) corresponds to scales of the order of the filter scale used to

define the environment, and hence halo masses must be restricted by the amount of matter that is available in their environment. We have taken this effect into account by restricting the Fourier modes that can contribute to the variance of the overdensity field in a given environment, suppressing those corresponding to scales larger than R_e. In practice we have implemented this by weighting each mode by the 'inverse' of the window function used to define the environment, W_e:

$$\sigma_{\rm eff}(M) = \int_0^\infty \frac{k^2 dk}{2\pi^2} \left[1 - W_e(kR_e)\right]^2 |W(kR_h)|^2 P_k. \tag{A.14}$$